Stacks Pool

NAVAL RESEARCH LOGISTICS QUARTERLY

OFFICE OF NAVAL RESEARCH

DEPOSITED BY U.S. Los Angeles Public Library JAN 3 1955

CONTENTS

ARTICLES

	Page
Navy Stratification and Fractionation for Improvement of Inventory Management *Rear Admiral F. L. Hetter*	75
A New Field for Logistics Research *Rear Admiral Bern Anderson (Retired)*	79
The Application of Statistical Techniques to Management of Overseas Supply Operations *E. F. Hollingshead*	82
Notes on Applied Analytical Logistics in the Navy *Captain Charles F. Lynch*	90
The Rommel Papers — A Commentary *Rear Admiral Henry E. Eccles (Retired)*	103
A Numerical Method to Determine Optimum Strategy *John von Neumann*	109
Optimal Weapon Systems *J. Marschak and M. R. Mickey*	116
On Some Applications of the Theory of Dynamic Programming to Logistics *Richard Bellman*	141
The Caterer Problem *Walter Jacobs*	154
NEWS AND MEMORANDA	166
RECENT PUBLICATIONS	169

VOL. 1, NO. 2 — JUNE, 1954

NAVEXOS P-1278

NAVAL RESEARCH LOGISTICS QUARTERLY

EDITORS

Rear Admiral H. E. Eccles, USN (Retired)
The George Washington University
Washington 7, D. C.

O. Morgenstern
Princeton University
Princeton, N. J.

Commander G. S. Essenwine, USN
Office of Naval Research
Washington 25, D. C.

F. D. Rigby
Office of Naval Research
Washington 25, D. C.

J. Laderman, *Managing Editor*
Office of Naval Research
Washington 25, D. C.

ASSOCIATE EDITORS

D. Blackwell, *Howard University*
R. M. Bowstrom, *Captain, SC, USN*
E. W. Cannon, *The George Washington University*
G. B. Dantzig, *The Rand Corporation*
M. M. Flood, *Columbia University*
J. D. Hayes, *Captain, USN*
E. B. Henry, Jr., *Commander, USN*
A. T. Magnell, *Captain, SC, USN*
J. Marschak, *University of Chicago*
T. C. Roberts, *Commander, SC, USN*
E. K. Scofield, *Commander, SC, USN*
E. D. Stanley, *Captain, SC, USN*
H. Strock, *Commander, SC, USN*
R. M. Thrall, *University of Michigan*
C. B. Tompkins, *National Bureau of Standa*
A. W. Tucker, *Princeton University*
R. E. Williams, *Commander, SC, USN*
M. A. Woodbury, *University of Pennsylvania*

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Manuscripts submitted for publication should be typewritten, double-spaced, and the author should retain a copy. Manuscripts and other items for publication should be sent to the Managing Editor at the above address. Authors are entitled to receive 50 free reprints of their published papers and may obtain additional copies at cost.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the months of March, June, September, and December and can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D. C.

The views and opinions expressed in the articles in this Quarterly are those of the authors and not necessarily those of the Office of Naval Research.

The printing of this publication was approved by the Director of the Bureau of the Budget, June 30, 1954.

NAVY STRATIFICATION AND FRACTIONATION FOR IMPROVEMENT OF INVENTORY MANAGEMENT

Rear Admiral F. L. Hetter
Supply Corps, U. S. Navy

World War II was largely a war of logistics. Faced with carrying on a global-wide conflict, the area and tactical commanders were required to rely on a responsive logistic supply flow which of necessity depended on intelligent planning for future needs and a guarantee that the goods would reach the combat areas in a constant uninterrupted flow. In an unprecedented scale, general use and technical items were delivered to highly mobile task forces and chains of advanced bases throughout the world.

To meet these demands a supply system, global-wide, amenable to proved supply management methods was developed and is in operation today.

This system is based on the principle of centralized control of policy and management, and decentralized operation, in both its material control and distribution phases. Centralized control of policy is exercised by the Chief of Naval Operations and the Assistant Secretary of the Navy (Chief of Naval Material). Centralized management of the material control and distribution phases is exercised by the Bureau of Supplies and Accounts. The operating controls are exercised by several supply demand control points designated "supply offices," each responsible for one or more specific categories of material and each reporting jointly to the Bureau of Supplies and Accounts and the technical bureau concerned. The actual materials handling and issuing are performed by activities in the field who report the status of their materials to the supply demand control points controlling the particular category of material.

It must be emphasized that this system developed by evolution. By continual scrutiny of its effectiveness notable achievements have been accomplished which have resulted in continual improvement of inventory management.

It is difficult, even for an experienced naval officer, to comprehend how much material was in the pipe-line when World War II terminated. It is believed it was Admiral Halsey who said, "You can't arrange battle plans so that the last Jap is shot with the last bullet." That illustrates the problem of pipeline residual material when a war is ended.

About the time the Navy was ready to dispose of World War II excesses, the Korean action forestalled the implementation of the program. It is now being pursued with vigor. We also found a need to change our supply management methods for the purposes of eliminating surpluses and of more effectively assigning major effort to significant items.

Almost two years ago the dollar value summary management reports of inventories, issues, and inter-activities movements, which are made to the Bureau of Supplies and Accounts, and to the Supply Demand Control Point Inventory Managers, were critically examined. The supply demands of the Korean action were being met and we had time for self-analysis. The management reports showed too small an issue figure for the inventories which were held, and, at the same time, inter-activities transfers were too high. We embarked on a further analysis of inventories and issues to use, on an item basis, which analysis is currently going on. The item analysis resulted in the accumulation of sufficient data with which to formulate a program of action. It is significant that the data reports and management controls which are built into the Navy Supply System were the tools without which our self-analysis, for the purpose of refining the system, would not have been possible.

Effective inventory management requires that inventories be separated into certain basic divisions which can be directly related to the purpose for which those portions of the inventory are held. Recognizing this precept the Navy initiated a stratification program. By establishing such a program, the desired result would be attained of providing more meaningful data for inventory management as well as for budget justification.

Certain necessary fundamental assumptions were made. One of these premises included the fact that mobilization reserves are held only at field activities in the stock status reporting system. These reporting activities are called distribution points or primary stock points and are Naval Supply Centers, Depots, Air Stations and similar major installations. Similarly, the line item computation of excesses would not cover material at non-reporting activities or secondary stock points such as a shop store within an activity. Non-reporting activities, generally, however, are required to report excesses to their supporting activities and are not authorized to carry mobilization reserves. Therefore, if the dollar value of their inventory, compared to issues, exceeds authorized operating stock levels, policing action can be directed by the Supply Demand Control Point to reduce inventories by issue, or transfer to their supporting activities as appropriate. Such a course of action would result in total money value of inventories reflecting system balances reported on financial reports submitted to the Bureau of Supplies and Accounts and complete coverage is assured.

It was determined that a breakdown of inventories that would accurately depict the inventories held in stores accounts should be segmented and encompass:

1. Excess, identified for disposition
2. Support for mobilization requirements
3. Personal property of a "capital" nature
4. Support for Mutual Defense Assistance Program operations
5. Support for current operations other than MDAP

The objectives of the program were firmed up and all inventory managers at the Supply Demand Control Points were directed to formulate the necessary mechanics for implementation. As a result, stratification was completed by 1 October 1953, although some of the inventory managers are refining their original stratification with another run.

Currently, the basic stratification of inventories is planned for accomplishment once every year at the central inventory control level.

The results of the first run have indicated the practical utility of stratification. For one, our inventories are no longer represented by a single unmanageable figure which requires elaborate explanation when requesting additional funds to augment existing inventory. It permits then a meaningful justification for fund requests. It encourages also more intelligent and critical consideration by all budget review authorities. This is true because funds requested are easily related to deficiencies in a specific strata—War Reserve, Special Program, or Operating Stocks. In other words, after stratification, the entire inventory then can be measured against that part of the inventory which should turn over—operating stock.

Since stratification of inventories involves a detailed analysis of items by programs, e.g., an analysis of the status of items on hand to support the remaining life of a plane type program, or our Mobilization Reserve position, it is valuable because it forces periodic detailed management review of inventory positions, it reveals balance or lack of balance among programs, and permits an evaluation of readiness at any given point and time, and an accompanying projection of readiness as of any specified future date.

Clearly, stratification also permits the isolation of excess stock for our own examination and subsequent disposal action. This enables us to set target objectives and police their accomplishments. It encourages frequent and periodic disposal action and cancellation of items still on order when they show up in the surplus or economic retention strata.

Even though stratification has proven itself valuable, we foresee much greater potential for refinement of the Navy Supply System from the companion analysis that inventory managers have been directed to accomplish—"Fractionation." Studies concluded by BuSandA indicated almost without exception that material inventories fell into definite demand patterns, with a relatively few items of those carried accounting for a very large percentage of demand. In view of these facts, it was considered necessary to classify immediately all material on the basis of demand patterns in order that maximum and enlightened inventory control effort might be placed on those items that do constitute most of the demand.

It was considered that quantity movement was the most important factor for consideration in fractionation since delivery and use of material is of primary importance in a military organization. However, dollar values of demand also necessarily must be of considerable importance since it is the responsibility of every inventory manager to control the Navy's dollar investment wisely. There, it is imperative that in any fractionation plan consideration be given to both velocity of movement and dollar value of demand.

If "Stratification" can be visualized as being accomplished by drawing horizontal lines across a block of inventory, "Fractionation" of the same block results in vertical lines from top to bottom of the block breaking the block into vertical columns of inventory. One column may be items having a high rate of issue with a high unit price; the next column may be composed of items having a high issue rate with a low unit price; the next may be items having practically no issue rate, a high unit price and such military importance that they must be maintained reasonably available for Navy use.

Now, of what management utility is fractionation? There are many significant uses—both in order to improve effectiveness and to improve economy of supply management.

The fractionation analysis has shown that in practically all categories of material, as low as 15 to 25% of the items will meet 90% or more of the demand. That fraction, which is high demand and may also be high in unit price, warrants, as a generality, more attention. We therefore can gain in both effectiveness and economy by assigning our best people in the Supply Demand Control Points to inventory review and provisioning of those items. We can also assure effective support for a statistical 90% of demand by positioning such items in all prospective points of use, at a relatively low cost in terms of storage, handling and record costs.

Similarly, we can assign best talent to high-cost fractions at provisioning review and replenishment. At the same time, it will be possible to eliminate regular central reporting on high cost insurance fractions and require reporting only on a transaction basis.

Fractionation also permits tailoring of the distribution system for economy with no substantial loss in effectiveness by stocking very low demand and insurance fractions, particularly those which are high cost, in only one or two stocking points and using premium transportation to meet demands at other points when they arise. It is significant that the Chief of Naval Operations has adopted this as policy. In effect, he has said that we are confronted with excesses in certain areas because we have tried to stock all items of possible use everywhere in order to be prepared for contingent demands. In effect, he stated that we can no longer afford to attempt to meet 100% of demand, immediately, everywhere.

In conclusion, the technology of our weapons is complex; our operational and logistics jobs are complex; the technology of our supply management is complex; it therefore demands tools to provide manageable segments.

The increasing understanding of this supply philosophy by Operational Commanders, Naval Engineers and Material Managers will provide our Navy the optimal support it so richly deserves.

Our keynote for the improvement of inventory management may be reduced to the very short title as enunciated by CNO, "Need Versus Desire." Stratifying the inventory and fractionating items within commodities are the most modern tools so far developed to accomplish this keynote.

* * *

A NEW FIELD FOR LOGISTICS RESEARCH

Rear Admiral Bern Anderson
U. S. Navy (Retired)

The more we penetrate into the general field of logistics research and seek the answers to the countless problems presented by that broad subject the closer it should lead to a study of similar problems that confronted the enemy. How he solved them, or failed to do so, is a profitable field for study in itself, for in it we may find lessons so important that they can point the way to changes in our own plans and methods. One theater of World War II, especially, appears to offer a fruitful field for research in that respect.

From the German side the campaign in Italy during World War II was an efficiently fought delaying action that upset all Allied time tables. It was always considered a secondary theater by the German high command and hence was apt to be on the slim side of the overall supply system. Yet in fighting its way up the leg of Italy in a series of well planned and executed delaying actions Kesselring's armies remained a formidable fighting force to the closing days of the war. They were able to be such from a high order of flexibility in tactical organization, an ability to meet new tactical situations with plans drawn up quickly in the field, and a supply system geared to the same tempo.

The first display of that ability to meet a critical situation rapidly and efficiently came in the closing days of the Sicilian campaign. From 10 to 17 August 1943 the German forces on Sicily were successfully evacuated across the Strait of Messina to the Italian mainland with all their equipment and without the loss of a man. Kesselring ordered the evacuation only the day before it began. Naval barges of the type built for the invasion of England, as well as a variety of other small craft, were used for ferrying the troops across the Strait, so some fast naval planning was also involved in the evacuation. The only Allied opposition to this movement was by the air forces. Whether or not more vigorous Allied action by surface naval forces could have broken up the movement is moot. From the German side, however, it was eminently successful.

When Montgomery's Eighth Army invaded the toe of Italy on 3 September only one German division faced it in a delaying action, in which skilfully used demolitions was a major factor, an extremely useful device in mountainous southern Italy which has few roads and railroads. Bridges were destroyed and railroads torn up. At the time of the Salerno landings on 9 September there were six German divisions south of Rome. One was in front of the Eighth Army, with one in the instep backing it up. One was near Naples, one was north of Naples and one was back of the Gulf of Salerno. An airborne division, with headquarters near Foggia, had elements dispersed over southern Italy. There was, therefore, but one division to oppose General Clark's Fifth Army on D-day at Salerno. Added to the Germans worries on that occasion was the task of disarming Italian troops after the Italian surrender announced only the evening before the landing.

Kesselring's reaction to the Salerno landing was prompt and vigorous. The reserve division facing Montgomery was ordered to Salerno and leading elements reached the battlefield in the evening of D-day. A Corps commander was ordered to move in and assume command of the defenders. The divisions near and north of Naples were also ordered into the battle. Later the division delaying Montgomery was told to break off the action and move to Salerno, and other

elements of less than division strength were sent to the beachhead. By 14 September, D-day plus four, the German forces at Salerno were built up to in excess of four division strength and were able to launch a major counter-attack against the beachhead. Obviously a smooth working supply and administration system had to be functioning in order to make those moves so quickly.

When the slow advance of Montgomery's army towards Salerno threatened the German positions there the German command organized a withdrawal in the form of a great arc pivoted on the seacoast south of Naples. The plan provided for taking up positions across the Volturno River north of Naples preparatory to taking up a "Winter Line," later made famous by its strongpoint at Cassino, being prepared across Italy and anchored to mountain strong points. The timing of the withdrawal coincided with General Clark's plans for a breakout—in a plan almost exactly complementary to the Germans. By the skillful use of demolitions and rear guard actions, the German time table was adhered to very closely. After extensive demolitions in Naples the last German troops evacuated the city the morning of the same day British troops entered it.

When the Allied armies came up against the Winter Line in November, they were stopped and despite many bloody attacks had only minor successes in breaking that line until May, 1944. Out of that stalemate grew the idea of the Anzio landing that took place in January 1944.

Until the middle of December 1943, there was a divided German command in Italy; Rommel in the industrial north and the Alpine passes and Kesselring in middle and southern Italy. At that time Kesselring was given command over all German forces in Italy, including navy and air forces, making him a true theater commander. He fully expected further Allied amphibious operations behind his front but the "wheres" and "whens" were unknown. The five most likely areas on the west and east coasts of Italy were chosen and detailed plans worked out to meet a landing at any of them. Each area was assigned a code word. Troops in northern Italy were designated to proceed to the landing site and elaborate movement schedules fixing routes, road or rail, timetables, and refueling points en route were worked out. For the Anzio landing on 22 January 1944 all that was necessary to start large forces converging on Anzio was the simple message "Situation Richard."

The speed and timing of the convergence of those forces on Anzio was remarkable. Kesselring always considered Rome as a logical Allied objective and normally kept two divisions in the vicinity in the event of an Allied landing near Anzio. But, linked to the landing, and a few days earlier, General Clark launched a major attack against the Winter Line. With it in danger of cracking Kesselring ordered the divisions near Rome towards the front, thereby taking a calculated risk as he believed the offensive was a preliminary to a landing, although he had no idea just where and when it would take place. As a result, when the Allies landed at Anzio there was only one battalion defending the coast from Anzio to the mouth of the Tiber and there was nothing to stop a dash from Anzio to Rome - except - the German potential in northern Italy.

Within a matter of hours German troops were moving towards Anzio. Battalions of infantry, tanks and artillery were pulled out of the main line along with regimental commanders and their staffs and formed into battle groups. Divisions en route to the main front were diverted. Divisions from northern Italy were set in motion. One light infantry division from southern France was in the Anzio area within ten days. And units were sent from Germany itself. An Army commander and his staff, normally stationed in northern Italy was functioning at the beachhead within five days. There were also two Corps commanders there.

Elements of German troops began to arrive in the Anzio area ready for action by evening of D-day and within 48 hours a strong perimeter was formed. By 29 January when General

Lucas had consolidated his beachhead and was ready to advance towards the Alban Hills only very limited success was had. The Germans launched three counter attacks in February. The main one in the middle of the month punched a salient in the Allied line but could not break through. Naval gunfire helped stop that attack. By then the Germans had elements of 15 divisions in the perimeter with large numbers of artillery pieces. There were well over 100,000 troops there, outnumbering the Allies. Another stalemate took place which lasted until May.

What makes this of major interest is that the concentration was made almost exactly as scheduled despite the fact that Allied aircraft dominated the air over Italy. The Air Forces made systematic raids aimed at interdicting road and rail traffic in the rear of the German fronts. Obviously, from the rapidity of assembly and the size of forces involved their effect was scarcely more than nuisance to the Germans.

Late in May 1944, the Allies broke the southern line, a breakout was made at Anzio and Rome was entered early in June. The Germans made another delaying action withdrawal to a new line in the northern Appennines guarding the Po valley. There too a long stand was made and the German armies in Italy were still a fighting force until April 1945 when, nearly a million strong, they collapsed and were the first major German forces to surrender.

This necessarily sketchy outline of the German campaign in Italy is given to suggest the remarkable flexibility and adaptability the German forces had, not only in the face of enemy ground forces but with enemy overwhelming air superiority. How could they do it? Italy was not blessed with many good roads and railroads, and that in itself created many problems. This is essentially an Army logistics problem, although the German Navy contributed materially to the supply by running barges and small ships along the Adriatic and the Ligurian and Tyrrhenian coasts behind protective minefields. Only small Allied craft, such as PTs, could get at that traffic but the German escorts were much more heavily armed than these light craft and there was little interference to the flow of supplies.

That remarkable German performance in Italy appears to offer a fruitful field for logistics research, from which invaluable lessons can be derived. What sort of supply organization existed that could plan for and supply such a diversified force as that which flowed into Anzio? What sort of planning factors did they use? How did their requirements conform with ours? What were their staff organizations and planning methods? How much material was sent from Germany and how much was obtained in Italy and the Balkans? What were the relations between the German Army supply system and the German Navy? In short, what is suggested here is an intensive study of the German logistics system both to learn why it worked so well and what we can learn from it. The campaign in one theater of operations can provide the basis for such a study.

* * *

THE APPLICATION OF STATISTICAL TECHNIQUES TO MANAGEMENT OF OVERSEAS SUPPLY OPERATIONS

E. F. Hollingshead
Naval Supply Center, Oakland, California

Supplying overseas naval forces is a complex operation involving many segments of the continental naval establishment and civilian economy. The isolation, analysis and evaluation of all factors in the supply cycle is a continuing problem and must be carefully planned and executed if deficient supply support is to be promptly detected, traced to its cause and intelligently corrected.

Statistical techniques have become important and growing management tools in overseas supply operations. The Pacific Requisition Control Office at the Naval Supply Center (NSC), Oakland, California, receives monthly tens of thousands of Pacific requisition line-item demands. In the Pacific Ocean Area there are several hundred requisitioning activities both ashore and afloat. With such a volume of requests for material the only way to glean facts from figures or to recognize significant trends is through an accumulation and analysis of organized statistical data.

More than half of the Pacific requisition line-item demands are filled from NSC Oakland stocks. Roughly 1,379,000 distinct items are included in the stock system. Of these 661,000 are carried by cognizant supply depots at NSC Oakland. Some items are fast moving, others have medium mobility or are slow moving. Some types of material are Bureau controlled. Some are non-standard items which are purchased. Many requests are passed to continental supply-demand control points for supply action.

The various NSC Oakland supply depots are faced with problems of stock control (inventory control of thousands of items with varying mobility); mechanization; financial edit; storage (including cold storage); issue control; supply action time.

The Freight Terminal Department (FTD) at NSC Oakland has the problem of inflow and outflow of material including receiving, consolidating, packing, documenting, shipping via various types of transport—air, parcel post, truck, railway express, freight, Military Sea Transport Service (MSTS) and Pacific Fleet Service Force vessels, as well as the constant goal of reducing time at tidewater.

The author's experience in applying statistical techniques to overseas supply operations has been with the Pacific Requisition Control Office (PRCO) at the Naval Supply Center, Oakland. Requisition control becomes a sizable problem requiring mechanization by means of electrical accounting machines, together with statistical analyses and reports. The following discussion is concerned with the statistical activities of PRCO.

The mission of PRCO is to coordinate efforts aimed at improving material logistic support to the Pacific fleet. The principal functions are:

1. Screening and channeling of Pacific requisition line-item demands;
2. Maintaining control of Pacific requisition line-items;
3. Maintaining records of material status by line-item;
4. Expediting and handling follow-up correspondence;
5. Closing out of the mechanized system requisition line-item demands which have been completed;

6. Ferreting out deficiencies in the Pacific supply system;
7. Providing analyses and reports aimed at pointing up deficiencies and evaluating improvements in material logistic support to Pacific;
8. Making primary estimates of tonnage of water cargo covering PRCO controlled materials for Military Sea Transport Service (MSTS) and Pacific Fleet Service Force vessels; and
9. Coordinating efforts of the various supply activities engaged in material supply to the Pacific Ocean Area.

Proper statistical control and evaluation provide management at various levels with guide lines on which to base policy decisions and to take action to improve operations. To mention only a few applications:

1. Singling out problem issues requiring executive action;
2. Evaluating supply performance segment by segment;
3. Analyzing the components of supply problems;
4. Taking action to correct deficiencies;
5. Setting up goals of accomplishment;
6. Providing operational motivation;
7. Improving supply procedures;
8. Providing average lead time for guidance of overseas requisitioning activities;
9. Reduction of overseas stock levels by reducing procurement lead time;
10. Obtaining an optimum balance between the effects of inventory control which brings material to demand and requisition control which brings demand to material;
11. Providing quality control;
12. Measuring productivity; and
13. Providing a factual basis for policy decisions.

The Pacific Requisition Control Office (PRCO) statistical reports and special studies are being utilized by management at various levels in taking action to improve Pacific supply support. Effort is being made by PRCO to keep these reports dynamic and alive. It has been found that reports that are firmed-up to the point where monthly figures are filled in to previously-designed formats soon lose their value as management tools. They become inflexible and unresponsive to new problems.

USE OF PRCO REPORTS AT NSC OAKLAND LEVEL

At the Naval Supply Center, Oakland, supply action time studies, as well as many other statistical analyses and reports furnished by the Pacific Requisition Control Office are used by the Commanding Officer, by the Planning Officer, by the Officers-in-Charge and the planning units of the supply depots, Material Services Department, and Freight Terminal Department.

By pinpointing the various segments of supply action time—issue time, consolidating and packing time, parcel post processing time, and tidewater time—supply performance may be appraised by management and steps may be taken to correct deficiencies or achieve further improvements. New goals of achievement may be determined. Thus an effective contribution may be made toward the solution of the total supply problem.

TIDEWATER TIME

A concerted effort is under way at the Naval Supply Center to improve tidewater time (elapsed days from date material is received at the freight terminal sheds to sailing date of vessel on which material is lifted). What is responsible for this impetus?

Recently the Commanding Officer studied the tidewater time charts prepared by the Pacific Requisition Control Office. Tremendous improvement during the past year was noted—the median average time at tidewater had been reduced from 21 days to 10 days. But the Commanding Officer felt that even further gains could be achieved. Noting the difference in tidewater time for Center-generated (issued from stocks at NSC, Oakland) and Off-Center-generated (passed to Off-Center supply activities for action), his expression reflected grave concern. Four—five—even six days difference! Wouldn't two days be more reasonable?

How did the Naval Supply Center go about resolving this problem of tidewater?

Targets of achievement were set up, offering both management and operational personnel a challenge.

1. Instead of moving 50% of the material at tidewater within 10 days, the goal was set at 75% within 10 days.
2. Instead of an average differential of 5 days in tidewater time between On-Center and Off-Center-generated material, the goal set was 2 days.

Freight Terminal Department (FTD) employees were invited to offer beneficial suggestions regarding changes in operational procedures aimed at reducing time at tidewater. (There is a program of Beneficial Suggestion Awards at NSC, Oakland). FTD personnel effervesced with good ideas. They cooperated. They coordinated their efforts. PRCO evaluated their performance and supplied charts so that FTD employees could see how they were doing.

Those involved in the responsibility for traffic operations felt strongly that the immense amount of paper work involved in the movement of fleet freight compared to parcel-post shipment could be substantially reduced through more liberal use of parcel-post privileges. For example, the following traffic clerk operations normally required for freight shipment are eliminated when parcel post is utilized:

1. Preparation of a packing list or recap;
2. Listing of the master invoice-number on all sub-invoices;
3. Preparation of a bill-of-lading;
4. Listing of the bill-of-lading number on all copies of invoice;
5. Stamping the shipping data on the issue-control copy of invoice (Under parcel post this information is incorporated in the packing date which normally is the date the packages are delivered to the Post Office);
6. Stripping the invoices and forwarding the accountable copies to the consignee under separate cover (In parcel-post shipments the accountable copies are included in the package along with the packing copies; no additional operation is involved);
7. Manifesting of the shipments; and
8. Checking of the manifest on the dock at destination—each item must be signed for as received.

Suppose parcel post is used and some items are lost. Can they be traced? Items shipped via fleet freight can be traced since they are covered by bills-of-lading and manifests. Not so with parcel post. Experience has shown, however, that it is a rare occurrence indeed for parcel-post packages to be lost. A few lost items now and then probably would cost far less than the additional manhours involved in terminal functions of fleet freight.

Is parcel post expensive in comparison to fleet freight? Not if items for the same destination are consolidated—within the parcel-post limits of weight and size of package.

To illustrate: Under the new policy the small routine items of Ordnance, Ships Parts and General Stores material are consolidated and the parcel-post shipment is forwarded weekly.

This practice has reduced the number of individual packages from approximately 50,000 a month to about 20,000, thus materially decreasing parcel-post cost. In addition to the saving in traffic work mentioned above there are additional savings in type of containers. A cardboard carton is used for parcel post; a more expensive wooden box is used for fleet-freight shipment.

But would parcel-post packages reach their destination as quickly as items shipped via fleet freight?

Much faster! Oakland Post Office trucks pick up parcel-post packages at NSC Oakland several times a day. Many more ships are available from the San Francisco Area for parcel post than for fleet freight. Some twenty vessels per month are reported available for mail shipments destined for the Far East, compared to two or three for fleet freight. Some of these vessels are fast-moving passenger ships. Besides, it is reported that tidewater time for parcel post averages about 4 days—considerably less than tidewater time for fleet freight.

The Freight Terminal Department decided to try diverting a larger number of small, routine-type items to parcel post. Many manhours normally involved in terminal functions for freight shipment were saved. Thus there was more time to devote to heavy and bulky items which must be shipped via fleet freight.

In a matter of weeks after the Commanding Officer suggested the new goals, time at tidewater for fleet freight over-reached the target! The goal was 10 days-or-less for 75% of the material at tidewater. The time was reduced to 9 days-or-less for 75% of the material! Moreover, the difference in time for Off-Center and On-Center-generated items was drastically reduced. From an average 5-day differential to 2! Just what the C.O. called for!

The above remarks refer to material destined for ships-and-mobile-units. Currently there is an all-out effort to make a similar reduction in time at tidewater for material destined for Pacific bases. FTD employees have made numerous beneficial suggestions. For example:

1. Booking material enroute from receiving shed to the packing line;
2. Cutting bill-of-lading while material is enroute from the packing line to the pier;
3. Consolidation of issues on outstanding obligations—placing about 9 released obligations on a single invoice, thereby reducing the cost of cutting invoices as well as the paper work involved in terminal functions.

MAJOR FACTORS OF TERMINAL FUNCTION

Some of the major factors of terminal function which contribute to tidewater time are:

1. Cargo handling, including receiving, segregation, marking and consolidating;
2. Documenting time, including billing and pre-planning;
3. Lead time required by Military Sea Transport Service (MSTS) for review and approval of proposed loading plan;
4. Time consumed in loading vessel;
5. "Surge time"—time loss when the amount of work exceeds the normal capacity (if the working complement were large enough to take care of the peaks, FTD would be over-staffed); and
6. Other factors such as ship cancellation, changes in on-berth dates, space allocation changes.

ISSUE TIME

Similar efforts are being made at the various supply depots to reduce issue time—elapsed days from receipt of requisition to issue of material. Formerly all items were issued on invoice.

Recently there has been a change in procedure for top-priority items with short lead time. Such material may be issued on a requisition if delivery of the material cannot await the preparation of a mechanically-cut invoice nor go through the usual steps of processing.

Prior to the introduction of the issue-on-requisition (IOR) procedure, high priority items averaged about 7 days processing time. Now it is typical for an emergency item to be issued on requisition within a few hours and go through packing within 4 hours.

Officers-in-Charge of the various supply depots watch closely the issue-time figures published monthly in the PRCO Requisition Control Analysis of Material Logistics Support to Pacific. If there is a worsening, immediate steps are taken to correct deficiencies.

ANALYSIS OF SHIPPED COMPLETIONS

The Pacific Requisition Control Office analyzed January 1954 shipments of items destined for the Pacific. Such demands were supplied as follows: 52% were issued from stocks at NSC Oakland; 6% were obligation releases; 42% resulted from shipments that were Bureau Controlled, passed to activities outside Oakland, or purchased on the civilian market.

"In Stock"

"In Stock" indicates that the line-items shipped were furnished from stocks at NSC Oakland.

Roughly 3/4 of such items shipped in January were requisitioned in December and January—a lapse of 1 to 60 days from receipt of request in PRCO.

About 1/3 of the Medical-Dental demands were requisitioned, supplied and shipped during the month of January; approximately 2/3 were requisitioned in December and shipped in January.

The percent of January shipments completing December and January demands were as follows: Medical-Dental 98%; Yards & Docks 88%; Ordnance 82%; Ships Parts 68%; General Stores 61%; Electronics 58%; average for all 74%.

"Obligated"

"Obligated" indicates that the requisition line-items were not in stock at the time the requisitions were received; the material was obligated by cognizant supply depots awaiting receipt of stock material. The shipments represent obligation releases when the material arrived for stock.

Line-item demands in roughly 1 out of 3 such cases were received from six months to a year prior to shipment; about 1 out of 7 were received over a year before.

It is readily seen that it takes far longer to obtain an obligated item than to obtain material from Oakland stocks. Such obligated material because of this delay is likely to engender considerable follow-up correspondence from the requesting activity. It is not surprising that the not-in-stock situation of a supply depot is a matter of concern to the customer and to supply management.

USE OF PRCO REPORTS AND ANALYSES AT PACIFIC LEVEL

Overseas, the fleet and force commanders find need for closer scrutiny of requisitioning practices of their ships and bases in maintaining reasonable supply discipline throughout the Pacific Ocean Area. Pacific requisitioning activities, both ashore and afloat, are provided with statistics which reflect the effect of misassignment of priority indicators and short or already-past destination delivery dates (DDDs) when the requisition is received at Oakland.

PRIORITY DESIGNATOR SITUATION

The priority indicator system is designed to provide means whereby supply activities may process requests and ship the most important requirements first.

Priority "A" Emergency is assigned requests for material necessary to place an inoperative vessel, craft, aircraft, shore station, or an essential installation thereof in an operative condition in order to carry out its assigned mission or schedule. Also included are materials required for health and hygiene, where non-receipt by the assigned destination delivery date would cause extreme discomfort or serious personal hazards. For priority definitions refer to paragraph 206, Pacific Requisitioning Guide, COMSERVPAC INST 4220.2B (or BuSandA Manual 23004).

To illustrate the careless interpretation of the above definition, "A" priority has been assigned the following items: staples, index cards, film, a drain plug for dentist's sink, a left front fender for a Ford sedan, a floor mat for a Chevrolet, decorative lettering for a Plymouth, as well as almost any item for which justification may be written as "inoperative equipment" without regard to the essentiality of the equipment—such as parts for a lawn mower.

Non-critical requests assigned priority "A" interfere with supply action on emergency requests, thereby defeating the purpose of the priority system. Assuming that unrealistic high priorities have received premium processing and expediting of documents and materials, such items are not eligible to receive premium transportation from the Naval Overseas Air Cargo Terminal nor from the Navy Central Freight Control Office, due to lack of sufficient justification as outlined in the Navy Shipping Guide, Article 1810.

Routine mail is sometimes used in forwarding "A" priority requisitions when dispatch or air mail transmittal would provide the issuing activity with precious additional days to meet the specified delivery date at destination. The average time to cover local screening of a requisition together with mail and transmittal time from Pacific Ocean Areas to PRCO, Oakland, varies from 9 to 13 days. Dispatch requisitions are received within a few hours.

PRCO commenced publishing its Priority Situation studies in the June 1953 issue of the publication referred to above. By June 1954 the ratio of "A" priority requisitions to total requisitions showed a drop of 64% for supply activities ashore in the Pacific Ocean Area and a drop of 28% for ships and mobile units, reflecting a tightening-up in priority "A" justifications.

DESTINATION DELIVERY DATE SITUATION

Pacific requisitioning activities designate on each requisition a Destination Delivery Date (DDD).

A sizable percentage of the Pacific requisitions passed to PRCO indicated DDDs already past or with only a few days remaining to cover issue time, consolidating and packing time, tidewater time, overseas shipment and off-loading time.

PRCO was concerned over the DDD situation since unrealistic delivery dates cause excessive follow-up action as well as unproductive expenditures of government funds to effect premium processing of documents and materials. When the material reaches the shipping phase the justification may not support the assignment of premium transportation as outlined in the Navy Shipping Guide. Assuming the justifications were acceptable sufficient premium overseas transportation would not be available to handle the volume of short DDDs.

Lack of understanding on the part of requisitioning activities of actual procurement lead time required may account in part for unrealistic DDDs. The primary cause, however, of past and impossible DDDs is the practice of Pacific bases and mobile units to assign DDDs based on delivery in the forward area.

To illustrate: The Pearl Harbor Naval Shipyard requisitions were assigned DDDs corresponding to the day the truck was driven to pick up Ordnance supplies at the Naval Supply Center (NSC Pearl) warehouse. If an item was not in stock, NSD Pearl passed the requisition to PRCO, Oakland, without changing the DDD.

PRCO channels NSC Pearl Harbor Ordnance requisitions to the Ordnance Supply Depot (OSD) at NSC Oakland. The Issue Control unit of OSD takes the DDD into consideration in determining whether to give high-priority handling (issuing on requisition as opposed to routine cutting of invoices). Since all were past-or-short-DDD-type requests there was a physical limitation as to which and how many could be given premium handling. There could be no assurance that the ones actually given premium processing were the ones that most required it.

To illustrate the interaction between unrealistic DDDs and issue-on-requisition processing at Oakland: In November 1953, 80% of the Hawaiian Area Ordnance requisitions passed to PRCO, Oakland, indicated DDDs that were already past. The DDD situation reported monthly by PRCO indicated a need for corrective action.

The Pearl Harbor Naval Shipyard recognized that the difficulties stemmed from shipyard requisitions. The problem was resolved either by (1) including a procurement DDD in addition to the issue-from-stock DDD on requisitions submitted by them, or (2) instructing the Ordnance Supply Depot, Pearl Harbor, to obtain by telephone from the shipyard a new advanced DDD which is placed on the requisition extracted to PRCO, Oakland. In appropriate cases procurement action is facilitated by use of dispatch extracts.

The PRCO January analysis of DDDs disclosed remarkable improvement in the Hawaiian Ordnance DDD situation. A mere 3%—instead of the November figure of 80%—assigned DDDs already past. The reduction in unrealistic DDDs set up an interaction at the Ordnance Supply Depot at NSC, Oakland: A corresponding decrease in premium-type processing from 75% of all top-priority requisitions in November to 15% in January. Sufficient lead time was allowed for processing high-priority demands without resorting to the more expensive issue-on-requisition procedure.

PROCUREMENT LEAD TIME STUDIES

The Pacific Requisition Control Office provides charts broken down by priority designator to indicate average procurement lead time required from various Pacific Ocean Areas to obtain requisition line-items from continental United States.

Procurement lead time must allow for:

1. Local screening in the forward area;
2. Mail and transmittal time;
3. Supply time, including screening and channeling, stock control, financial edit, issue of material;
4. Consolidating, preserving and packing;
5. Terminal functions and awaiting shipment at tidewater;
6. Overseas shipment and off-loading time.

These charts are included as Encl. (1) in the Pacific Requisitioning Guide, COMSERVPAC INST 4220.2B.

USE OF PRCO REPORTS AT CONTINENTAL SUPPLY-DEMAND CONTROL POINT LEVEL

At continental supply-demand control point level, management is informed of NSC Oakland supply responsiveness to Pacific requisition line-item demands processed through PRCO.

Pacific demands have already been screened in the forward area—are harder nuts to crack. For this reason a supply depot's not-in-stock position will be worse than for continental demands.

In November 1953, PRCO made a study of the "In Stock" and the "Not in Stock" situation at NSC Oakland, covering system stock items only. In January the study was repeated. It was revealed that the "In Stock" situation had improved an over-all 5%.

USE OF PRCO REPORTS AT BUREAU LEVEL

At Bureau level PRCO analyses and reports are given consideration in gaining perspective over the entire Pacific supply situation—are used as guides to policy formulation.

MEDICAL SUPPLY SUPPORT TEST

The Munitions Board Medical Supply Support Test began 1 March 1952; support by the Army's Alameda Medical Depot was extended indefinitely by ALPACFLT 64—1952. Since March 1952 Medical-Dental requisition line-items have been packed at the Alameda Medical Supply Depot. Packed items are trucked from Alameda to tidewater at the Naval Supply Center, Oakland.

Until a statistical study was made by PRCO it was assumed that trucking required 1 day-or-less. But a pilot study in June 1953 revealed that the range was from 1-to-10 days; that 50% of the items consumed up to 6 days. Time from packing date to sailing date of vessel not only included time at tidewater but trucking time as well.

The study has been repeated monthly and published in the PRCO report referred to above. Results disclose that the median average trucking time has decreased from 6 days to 1 day. The range has contracted from 1-to-10 days to 1-to-5 days.

SUMMARY

The Pacific Requisition Control Office has demonstrated that statistical techniques are helpful to management as tools in appraising the Pacific supply situation, in correcting deficiencies, in setting up goals of accomplishment, in changing to more efficient procedures, in basing policy decisions on facts derived from figures—in short, in controlling supply action in its various segments.

* * *

NOTES ON APPLIED ANALYTICAL LOGISTICS IN THE NAVY

Captain Charles F. Lynch
Dental Corps, U. S. Navy

The title of this article should be tautological. Unfortunately, it is not at present. Despite recent advances in logistics theory and practice, and aroused interest in support as distinguished from operating functions, the logistics field is so wide and its ramifications are so many that there still remains much that is intuitive, empirical, or capricious and unscientific; that is to say, not based on or substantiated by analytical methodology. Several areas in which this is so come immediately to mind; for example, consumer stock levels, local purchase, standardization, speedier deliveries, and strict central monetary inventory control. Something about the first, and such allusions to the others as seem indicated because of their effect on or relation to stock levels, will be advanced here from the standpoint of a practical observer. This may be of some interest to the practical operator, but not to the theoretical worker, as precision will be slighted for the sake of emphasizing principles and approaches. In that sense, perhaps the title is misleading. "Semi-analytical logistics" might better describe what follows. If so, exculpation is claimed on the basis of Lotze's observation on the need to put ontology to use. He said, concerning faddish pre-occupation of his contemporaries with the character of knowledge to the exclusion of its application, "Continued sharpening of the knife becomes tiresome if we cut nothing with it."

It need not necessarily follow that because something is done intuitively it is ipso facto incorrect. As a matter of fact, when so done by persons with practical experience and technical knowledge, results usually are reasonably good and often are not too far from the best attainable. For example, prior to World War II, the minimum and maximum quantities of Bureau of Medicine and Surgery supplies at consumer activities were quantities equal to one and two years' use respectively; except, of course, for material subject to deterioration. "What an inventory and what a waste of money!" would be the present day comment. "How much more economical are the new levels!" (1/2 and 2 months, respectively) might be added. The old levels were not arrived at scientifically. They did reflect experience, however, because they were prescribed by persons who had been storing, using and ordering that material for years. Empirical and intuitive they were, it is true; capricious they were not. Incidentally, the old terms, "Minimum Quantity" and "Maximum Quantity" correspond to the present day terms "Safety Level" and "Stockage Objective." Present stock levels are the result of drastic reductions. Below table indicates the trend:

	Safety Level	Operating Level	Stockage Objective	Order & Shipping
Prior WW II	12	12	24	Local experience
Post WW II	3	3	6	1.5
Current	0.5	1.5	2	1

A pause and a look, not a detailed scientific analysis, but an examination based on practical factors, might reveal that the oldtimers, like the happy bachelor in the song, were "not so very far wrong" at that. A few of these factors follow:

a. Bulk. - Units are so small that whether a 1-1/2 or 12 months stock is carried makes little difference, so far as storage is concerned. Except for hospitals and quite large stations, storage room is negligible.

b. Packaging. - Unit packages, although small physically, contain such large quantities, in terms of usage, that the smallest unit often is enough for several months consumption in small and medium sized stations. In these cases low levels are an impossibility.

c. Erratic Usage. - Usage is so unpredictable in some cases that there is an unacceptable risk of running out with a safety margin less than three or four times the Order and Shipping Time usage.

d. Consumer Workload. - Each time stores are ordered, and especially when they are received, a good deal of time-consuming paper work has to be done. Similarly, considerable work is involved in uncrating, and stocks must be rearranged in order to adhere to FIFO (first in, first out) practices (incidentally, the FIFO system has been intuitively accepted where taxes are not a problem but, the LIFO (last in, first out) system may, by analytical methods, be shown to be meritorious, even in a Government agency, for material of certain characteristics). Since the workload for large orders is not proportionately greater than for small orders, fewer and larger ones are preferred.

e. Depot Workload. - Depot workload and, therefore, the number of employees are decreased for similar reasons and also because emergency orders and the percentage of repacks are lessened by larger consumer stock levels.

f. Transportation Rates. - Distribution of this particular material is like a mail order business. Hence, transportation costs cannot be reduced much but the larger levels would occasionally enable a slight saving because of more carload shipments.

g. Cost of Material. - The low cost of material in terms of dollars per person per period of time tends to deprecate the importance of capital investment represented by large inventories.

h. Clotty Issue. - 10% of the items comprise 90% of the total issues. This is characteristic of medical and dental supplies. For example, 15% of Parke Davis products (drugs and biologicals) account for 85% of sales.[1] This issue pattern emphasizes the impracticality of small levels for 90% of the items.

i. Obsolescence. - Characteristics have received close scrutiny with respect to central procurement and depot levels but not at consumer level where lead time is short and no purchase commitments are made. The incidence is very high. As an indication, 90% of Parke Davis products currently being sold were non-existent in 1930.[1]

All these, except the obsolescence factor, support a policy of few orders and large stock levels for this kind of material and other kinds with those characteristics.

But large stock levels are distinctly out of current favor. It is perhaps significant that the former large levels were set by those who worked at consumer, depot, and departmental level; whereas the current small levels resulted from extra-Navy budgetary decisions demanding conformance as a condition to release of funds. Obviously, the major areas where small levels are potentially economical are those of capital investment and cost of storage space. Secondarily,

---

[1] Corporate Director, October 1953, Audit #114, Am. Inst. of Mngt.

there is less chance of deterioration and obsolescence of material. Since conditioning is a powerful influence in human response, it is not surprising that finance and budget personnel tend to favor and impose small levels while operating personnel tend to favor liberal ones.

The new smaller levels, like the former larger ones, are intuitive but with different weight being given by different persons to the factors involved. They are empirical only in the sense that for other items, indeed for most other categories, the more rapid the turnover the better and, by implication, the better it is to have small inventories - the smaller the better regardless. They may be capricious - time will tell.

The question is can the different views be resolved without waiting. Before the work of scientific personnel was made available to the Navy by logistics seminars, enrollment in business management courses, Office of Naval Research-sponsored projects, etc., this would have been impossible. With the benefits of such cooperation and participation, there is reason to believe this matter of stock levels is capable of analytical resolution.

In this article it is proposed to attack the problem in this way. First, select a common item, as a point of reference, and ascertain what its optimum consumer stock levels are under varying circumstances from the standpoint of economy solely. Second, having found the most economical levels, test them for feasibility and acceptability in other respects, especially as to risk of depletion. Third, derive from the reference item data, ideas or approaches to a means of setting up standards for stock levels that will meet practical requirements; i.e., that can be executed easily in the field and yet adhere adequately to the requirements found by analysis.

Of the category under consideration, one of the most commonly and constantly used items is aspirin. It is suitable as a reference item because it is familiar and other items, or classes of items, with significantly different characteristics can be readily related to it in terms of what adjustments are necessary for similar analyses.

Obviously, the higher the item cost, the larger the item, the greater the use, and the faster the deterioration or obsolescence; the more profitable it is to have the lowest possible stock levels, even at the expense of increased administrative costs. The converse favors higher levels. Hence, there should be a point where it costs more to either:

Have a higher level and order less often

or

Have a lower level and order more often.

That point can be found by iteration somewhat as follows:

a. Annual inventory carrying charge = 4.5% of inventory, viz:
   164 units (M's) = $100.00 = annual usage 5200 personnel[2]
   1 sq. ft. storage space required for 164 units[3]
   $1.50 annual cost per sq. ft. storage
   $3.00 interest on inventory capitalization (3%)
   ∴ 1 x 1.50 + 3 = $4.50 per $100 inventory = 4.5%
b. Assume overhead cost each requisition and delivery = .60
   .40 depot cost per line item (cost accounting records)
   .20 consumer cost per line item (estimated)

---

[2] Recurring issue rate FY50 - 31.3/1000men/year.
[3] 1.5 cu. ft. per 24 units, price $14.64.

Then the total overhead for a one month stock level would be:

$\frac{1}{2}(100/12) \times 4.5\% = 0.19$ annual inventory overhead

$12 \times .60 = 7.20$ annual overhead for orders

7.39 total annual overhead

At this point, it should be mentioned that although the above is the conventional computation, it assumes application of the inventory carrying charge to only half the stock level; i.e., the average inventory. This is not entirely applicable to a rigid requisitioning cycle whereby a Naval activity periodically submits one requisition with many line items. For example, if items, a, b, c, etc., are ordered individually, then when a is received, b, c, etc., will be partially depleted and the vacated storage space can be used; but if a, b, c, etc., are ordered simultaneously, space must be available for, and that part of the carrying charge that reflects cost of storage space must be applied to, the full stock level. Since the Navy, depending on circumstances, uses either system of requisitioning it seems well for the sake of simplicity to adhere to the conventional approach with the mental reservation that it applies to some material but not to all and to some (very large) stations but not to others. Similar computations would show other stock levels to cost as follows:

Stock Level	Annual Inventory Carrying Costs	Annual Ordering Costs	Annual Overhead
2 months	.38	3.60	3.98
3 months	.56	2.40	2.96
4 months	.75	1.80	2.55
5 months	.94	1.44	2.38
6 months	1.12	1.20	2.32
7 months	1.31	1.02	2.33
8 months	1.50	0.90	2.40
9 months	1.69	0.79	2.48
12 months	2.13	0.60	2.73

This indicates that 3 months is too small and 12 months too large for this particular item with this amount of usage. It would appear that a level of close to 6 months would be theoretically ideal and that levels from 5 to 8 months inclusive would be practical since the differences in that area are not great.

Such computations are rather cumbersome for everyday use. Fortunately, mathematicians interested in inventory problems offer business a simple formula which can be readily modified to indicate a more useful Navy figure; namely, quantity in terms of months usage. This formula (with X = most economical order quantity in dollars) is:

$$X = \sqrt{\frac{2 \times \text{Value Annual Use} \times \text{Cost per Order}}{\text{Inventory Carrying Cost in \% Inventory.}}}$$

Then the desired formula can be developed (with Y = most economical order quantity in months usage) as follows:

$$Y = \frac{12\sqrt{\frac{2 \times \text{Value Annual Use} \times \text{Cost per Order}}{\text{Inventory Carrying Cost in \% Inventory}}}}{\text{Value of Annual Use}}$$

Thus, using values assumed in the iterative computation:

$$Y = \frac{12\sqrt{\frac{2 \times 100 \times .60}{4.5\%}}}{100} = 6.19 \text{ months.}$$

In round numbers then, a 6 months supply is the best amount to order from the standpoint of economy when there is a nonfluctuating demand. It is worthy of note that if in this example the annual use were increased to $200 or enough for a station of 10,400 personnel, 4.38 months would be optimum. By the same token, if an item of lower usage but comparable bulk and unit cost were substituted (which is the classification in which by far the majority of items in this category fall) much higher levels would be optimum. In fact, they would equal or exceed the pre-World War II levels. In other words, theoretically and perhaps practically, large stations should have lower stock levels than small ones and items with high usage or value should have smaller levels than those with low usage and unit value.

It is realized that in both iterative and formulative approaches the "cost per order" does not include transportation cost. In the former approach it could, of course, be easily factored in as simply (cost to transport year's use) / (number of orders per year). However, since the transportation cost per order, unlike the line item paperwork and handling cost, varies with the size of the order, inclusion would complicate the formula. Exclusion does not invalidate the equation because the annual transportation cost would be almost the same whether many small or few large shipments were made.

Next should be considered extension of the benefits of analysis of this particular aspect, the best quantity to order, to other stock level features. In this connection, it is interesting to observe that one of the fairly recent changes which encouraged a more scientific approach to the solution of logistic problems in the field of stock levels, was a change which at first glance would appear to be merely a matter of nomenclature, but which has a greater significance. It is the definitive breaking down of stock levels to two increments - Safety Level and Operating Level - the former being the old Minimum Quantity, the latter being the amount used during the requisitioning cycle, and their sum, the Stockage Objective, being the old Maximum Quantity. The significant point here, to be utilized later in extending the analysis, is that Operating Level, Requisitioning Cycle or Replenishment Period, and Amount to Order are all the same in terms of months usage. To repeat, at the risk of expanding a grasp of the obvious to a stranglehold, if one orders every 3 months he orders a 3 months supply and, if his best order amount is something other than 3 months then so is his Operating Level and Requisitioning Cycle. This pertains only to consumer "one shot" requisitioning of standard material; not to central procurement because obviously any desired Operating Level could be obtained by ordering on whatever cycle was convenient and scheduling deliveries appropriately.

If, instead of being an unknown variable or an arbitrarily-arrived-at constant, the optimum Operating Level is determined scientifically, it should be possible, so armed, to explore two important intuitively evident conclusions with regard to the Safety Level, the other element of stock levels; namely:

a. The larger the usage in a period, the smaller proportionately the Safety Level need be to provide a uniform probability of depletion.

b. The shorter the period between requisitioning and delivery, the smaller the Safety Level may be for a given probability of depletion.

With respect to the first point, the proportion of deviations from mean usage generally decreases as usage increases except in the case of variations resultant to obsolescence, to use

of optional items and to similar purely technical circumstances. If one breaks a bottle of acid it means little if 100 are used in a period but much if only two are used. Deviation from chance is apt to be small if a coin is flipped 10,000 times; large if flipped 10 times. Categories differ tremendously with respect to whether purely "technical" deviations, for want of a better word, from mean usage are important or whether chance deviations are of principal concern. The significance of and what to do about the former deviations can be assessed only by those who have specialized technical knowledge. But the latter type is amenable to a somewhat more analytical approach.

If the expected deviation from mean usage in % is accepted as conforming to $K \frac{\sqrt{\text{mean usage}}}{\text{mean usage}}$, it is possible to arrive at a theoretical safety allowance for any Operating Level. Then, by adding to that safety allowance a similar safety allowance for the Order and Shipping Time, a theoretical Safety Level adequate for chance variations in usage, but not for delayed shipments or "technical variations," could be derived. To refer again to aspirin, K = 1.26. Using that value and the previously illustrated best Operating Levels for stations of 10,400, 5,200, and 2,600 personnel and assuming a one month Shipping Time, results are:

Personnel	Months Safety Allowance	No. Units
10,400	.55	15
5,200	.9	12
2,600	1.5	10

Whether any appreciable number of items used by the Navy can be found to be possible of such treatment would depend upon statistical investigation. Although used for illustrative purposes, this item, one of the most widely and constantly used in its category, cannot be handled this way. One of the parameters not reflected is the influence of climate. In the Northern Hemisphere, usage is higher in December than July. To require factoring such data, would impose a burden on the field which it could not discharge. The inclination is to relinquish, in this instance, the possible exploitation of a decrease in variations with an increase in usage because of the presence of other influences of much higher magnitude. For the category used here for illustrative convenience, such work as has so far been done indicates the number of items to be so small as to contraindicate practical use of this approach. Obviously, when 10% of total items account for 90% of issues there can be only a narrow band of items with any possibilities along this line. As a matter of fact and of some possible interest by way of comment, a recent survey of six of the largest users of this category revealed that of the 6,000 depot items there were 3,000 that were not issued to use at all during a six month period, and 1,300 that were issued only once a year. Of course, many of them were nonconsumables and much more investigative work must be done, but the rather obvious step is to analyze the suitability and feasibility of requiring local purchase of such items instead of central procurement and depot issue.

Local purchase as an optional form of procurement is an interesting proposition, so a very brief consideration of one of its facets may warrant diverging a moment from the matter of safety allowances. Since an example is helpful if its features can be translated, so to speak, into terms of one's own experience, it is perhaps useful to consider an article exhibiting characteristics shared by other articles; which characteristics suggest local purchase.

Take a unit which is an assortment of items that could be obtained individually from local commercial sources, but which for purposes of convenience, or other reasons, are combined in one package for issue from a Navy depot. This situation is similar in one respect to the broader and also interesting matter of automatic supply and its handmaiden, block loads, in that it is impossible to devise an assortment of a large variety of items so that when one is exhausted all will be consumed. Inevitably there will be some items (or sizes of items if the assortment consists of a single item) left over. As successive assortments are received and consumed, the excess accumulates. The question is should this excess and ultimate probable scrapping of excesses be accepted or should local purchase be used to supplement depot shipments in such a way as to prevent development of these excesses. Obviously, the answer depends upon the relationship between cost per item by central purchase, cost by local purchase, and proportion of excess if assortments only are procured. Implicit in the proposition is the assumption that local supplementary purchase will prevent development of excesses.

The following formula is a means of expressing and evaluating that relationship. For the sake of simplicity, certain parameters such as effect of return by taxes, depot cost to hold, etc., are omitted.

Let X = % items in assortment becoming excess
Y = % usable items in assortment (100% - X)
Z = no. items received in given period by an activity

when the only supply source is the depot.

Then XZ = no. items received but becoming excess
And Z-XZ = no. items needed in given period by an activity
Also, let a = average cost per item depot procurement by assortments
b = average cost per item local procurement by item
p = total cost if depot is only source of procurement
q = total cost if supplementary local procurement used

Assuming that the prices of the individual items in the assortment do not vary much, and X and Y still hold approximately when local supplementary purchasing is used, then approximately

$$\frac{p}{q} = \frac{Za}{(Z-XZ)Ya + (Z-XZ)Xb}$$

EXAMPLE (Taken from actual experience)

X = 20%
Z = 16660 (10 Assortments; SN 5-606-168)
a = .80
b = 1.20

Then

$$\frac{p}{q} = \frac{13{,}328.00}{11{,}728.64}.$$

Thus, under above conditions, a reduction in cost of about 13% results when depot procurement of assortments is supplemented by local procurement.

A convenient table derived from the above and using X as a reference point, is as follows:

If X equals	Then break-even point is when
5%	$\frac{a}{b} = \frac{1}{2.03}$
10%	$\frac{a}{b} = \frac{1}{2.11}$
15%	$\frac{a}{b} = \frac{1}{2.18}$
20%	$\frac{a}{b} = \frac{1}{2.25}$

A value of a relative to b lower than shown would indicate it is cheaper to throw away unusable excesses than to attempt to keep inventory in balance by local purchase. A higher ratio indicates the converse.

To return to the matter of safety allowances, it appears, with respect to the second point, (the shorter the Order and Shipping time the smaller the Safety Level) that the situation is clearer and more significant from the practical point of view and highlights an area in which improvement, if possible of accomplishment, would effect real economies in inventory without imperiling safety of supply. This is so because, assuming the best Operating Level, which by definition is the best order quantity, has been established, the Safety Level which is an additional amount should be kept as small as possible. There are no opposing considerations such as cost to order, etc., as in the case of Operating Level.

For example, with a one month Safety Level a consumer with a 1-1/2 months Order and Shipping Time would run completely out if he used 2-1/2 months stock in 1-1/2 months; i.e., 167% of the norm. But, if it were possible to reduce the Order and Shipping Time to two weeks, the consumer would run out only if he used 1-1/2 months stock in 1/2 month or 300% of the normal usage.

Wolf and Kruskal[4] have shown the probability (assuming standard deviation equal to $\sqrt{\bar{x}}$) of depletion with a 3 months Safety Level and monthly usage of .182 to be as follows for various Order and Shipping times:

.5 months 3.5%
1.0 months 10.0%
1.5 months 14.9%
2.0 months 18.3%

Although above differences in probability of depletion are possibly atypical because the usage quantity is low - a usage of 1 per month would show a much smaller range - they focus attention on the need to reduce the time interval between ordering and receiving material. These investigators further showed that to achieve a uniform probability of depletion of say 10%, Safety Level,

[4] "Protection from Depletion of Supply in the Bureau of Medicine and Surgery;" J. J. Wolf and J. B. Kruskal, Issue No. II of Logistics Papers of the George Washington University Logistics Research Project, May 1950.

where actual Order and Shipping Time is something other than 1 month, should be reduced or increased to below quantities:

Months Ord. & Ship.	Months Safety Level
.5	2.1
1.5	3.7
2.0	4.2

Associated with this observation is the natural question as to whether it is worth-while not only to pursue every means to shorten Order and Shipping Time that can be accomplished without added expense but also even to pay a premium; such as air shipment, express, or motor haul rates instead of freight or parcel post rates. It would appear that one principal consideration is involved - cost of higher shipping rates vs. saving in inventory costs by reason of reduced Safety Level. Obviously, for highly valuable material or for material whose storage for one reason or another is expensive, it would be advantageous to accept the higher shipping rates. For example, if a unit of G-3 gold (a relatively low carat type with high Navy use), price $3.00, were substituted for aspirin in the examples of costs by iteration given earlier, and if usual shipping rates were .20 per pound from depot to consumer and air mail rates were .70, difference in annual cost incident to reduction would be as follows:[5]

Ord. & Ship. Reduced	Increase in Trans. Costs	Decrease in Inventory Costs	Net Decrease
.5 months	1.30	1.83	.53
1.0 months	1.30	3.65	2.35
1.5 months	1.30	5.48	4.18

In other words, theoretically and perhaps practically, constant effort, including special shipment at premium rates of high value items, should be exerted to lower Ordering and Shipping Time.

The matter of stock levels is not a completely isolated self-important topic. Final decisions must give weight to related considerations. While they cannot be gone into in this article that would require careful, extensive exploration - it perhaps is well to note in passing that such things as stock fund operations at consumer level and simplification and standardization, to mention only two, are part of a complex where each element influences the others. For example, the Committee on Government Operations (Riehlman Subcommittee) and others have recommended extension of stock fund operations to station level. The objectives sought are monetary control of stocks and inclusion of station stocks in the recorded over-all Navy inventory. Regardless of the manner of accomplishing these objectives - and there are simpler ways for example, a periodic report showing dollar value of issues to use and stock balance which may easily be made since the Navy, contrary to Army practice, prices depot issue documents - implementation affects stock levels and vice versa. Monetary control might be expected to take this form: "Your usage last three quarters totaled $9000.00; your inventory is $6000.00; prescribed Requisitioning Objective is three months; therefore, your 4th quarter allotment is reduced by $3000.00." It is easy to see that low stock levels might permit such action if the

[5] Recurring issue rate FY50 - 63/1000men/year.

inventory items were uniform as to unit demand and cost. But what of material where a large part of the items are in units exceeding prescribed levels? If the whole is to be held to a single stock level on a monetary basis then the most used items must be at something less than the over-all level; a situation which could be serious with respect both to danger of depletion of the most important items and to generation of too many emergency requisitions. One answer might be simply to specify more sizes of units.

That brings up simplification and standardization, an area of real pay dirt. An exaggerated and unrealistic but illustrative example would be an item with ten different lengths, widths, thicknesses, and colors. A stock of one each would be 10^4 or 10,000 items. If only one color were specified, 10^3 or 1000 items would be required. If then only one thickness were specified, 10^2 or 100 would suffice. It appears then, that one should, before applying this type of monetary control, weigh the merits of increasing stock levels against increasing the variety of sizes or doing neither and, depending upon the material, size of station, and other applicable factors, allow for more than the bare difference between value of inventory and predicted value of usage.

To return to stock levels per se, it seems that the possibility of making practical application of scientifically indicated measures can be accomplished by action somewhat as follows:

a. Each bureau categorize the material in its various classes in as few groups as possible based upon their relative annual inventory carrying costs in percent of inventory. As noted earlier, the elements concerned are: annual cost to store $1.00 worth plus interest per $1.00 of capitalization (.03). In the example given - aspirin - annual inventory carrying cost was about 4.5% based on a gross cube of 1.50 cubic feet for 24 units valued at $14.64. Extremes in the Bureau of Medicine and Surgery category are gold costing about 3% and plaster costing about 12%. Derivation of cost-to-store figures can be accomplished by EAM methods for all items as punched cards showing units per shipping carton, cube, and price are on hand. All that needs to be done is to: (1) divide cube by price; (2) multiply the quotient by an estimated annual rental cost per cubic foot; (3) add 3¢. Then, by dividing the range into three segments giving an appropriate arbitrary value to each segment (say 4%, 8%, 11%) designate items A, B, or C as they fit and include these designations in catalog item descriptions for guidance of consumers.

b. Similarly designate and catalog items as say I, II, and III on the basis of value of annual use per 1000 personnel. Values may readily be found by EAM (electric accounting machine) methods from records of annual recurring issues from depots. To again refer to the aspirin and G-3 gold examples, recurring issues in FY50 happened to be $19.00 and $185.00 respectively per 1000 personnel.[6] Extremes in the entire category are about $.02 and $2000.00 per thousand men per year.[7] This range would be too great to handle except that 90% of the items happen to be in the lower part of the range. By selecting out a few items such as plasma, serum-albumen, gold, and aureomycin for individual treatment, a range from about 15¢ to $40.00 covers virtually all items.[8] Segment values might be $1.00, $10.00, and $30.00 for items designated I, II, and III. There is not much use in going below $1.00 (even though 85% of

---

[6] Recurring issues FY50; aspirin (M's) $9300; G-3 gold $91,000.

[7] Recurring issues FY50; arsenic pentoxide $11; aureomycin $975,000.

[8] A random sample consisting of 20% of all consumable items indicated values of total annual recurring issues per 1000 men FY50 to be:

Under $1.00 for 85% of the items.
Under $5.00 for 91% of the items.
Over $30.00 for 3% of the items.

the items actually do fall below that figure) because, as will become apparent later, the theoretically best Operating Level would then be so high, even for a C item in a large station, as to be obviously impractical from the standpoint of obsolescence and deterioration.

As a very rough but practical expression, the square root of the ratio of b. amount to a. amount adjusted to express months usage instead of dollar value, would indicate theoretical, reasonably proper Operating Levels (not giving effect to technical considerations, such as obsolescence, deterioration, etc.) for a station of 1000 personnel. For an A III item, the expression would be $\frac{12\sqrt{30/.04}}{30.}$. Similarly, a table for all values might readily be compiled and extended to include stations of 5,000, 10,000, and 15,000 or more population; viz:

Item Designation	1,000	5,000	10,000	15,000
A I	60	26	19	13
A II	19	9	6	5
A III	11	5	3	3
B I	42	19	13	11
B II	13	6	4	4
B III	8	4	2	2
C I	36	16	11	9
C II	11	5	4	3
C III	7	3	2	2

This table may be condensed to an administratively convenient guide for consumer activities somewhat as follows:

a. Operating Level for stations of 2,500 personnel or less is 12 months regardless of item.

b. Operating Level for all items designated "A I," "B I," or "C I" is 12 months regardless of size of station.

c. Other Operating Levels in months are as shown in the table below:

Item Designation	Personnel Supported	
	2,500 to 7,500	7,500 or more
A II	9	6
B II	6	3
C II	6	3
A III	6	3
B III	3	3
C III	3	3

Thus, any consumer can record the proper Operating Level for each item very easily by referring to the catalog and guide.

In passing, it is interesting to note that the levels of 85% of the items approximate the old empirical level of 12-24 months. Remembering the size of the Navy and of Naval Stations prior to World War II, one is compelled to concede that the old fogies probably were not such bad managers at that.

Earlier it was stated that, optimum Operating Levels having been determined, Safety Levels should be amenable to analysis with a view to reduction to the lowest quantities; there being no contravening factors. For many materials it would seem that here is an area possessing more possibilities for economy than Operating Levels where after all, any decrease is accompanied by an increase in administrative costs which may approach, equal, or exceed the decrease in inventory costs. It is also a much more risky area because while it is quite easy to contend with unexpected heavy usage, while a considerable part of the Operating Level is on hand, by use of Graphomatic signals or other devices to indicate the Emergency Order Point, it is a different matter after a routine replenishment requisition has been submitted. Not only is the danger of depletion of concern but so also is that of excessive incidence of emergency orders. They, by their disruptive influence on routine shipments and high processing cost, very quickly nullify economies attained by low inventories. Not only that but a few depletions will make a pack rat out of almost any consumer and "issued stores," "shop stores," "ward stores," and other euphemistically named black market stocks that do not appear on any inventory report or stock account but which really are part of the Safety Level will soon accumulate; this to the disadvantage of a sound system and notwithstanding inspections.

In the category used here for illustrative purposes, depletion theoretically occurs when the consumer used 1-1/2 months stock in 1 month or 150% of the average usage. In reality, the Order and Shipping Time of activities near a depot is shorter than that of distant activities. Furthermore, it varies with the kind of material and for different sized shipments of the same material since a choice as to mail, express, railway freight or trucking depends upon those factors. The latter factor, in turn, depends upon size of consumer activity and size of Operating Level. The question is can these parameters be satisfactorily exploited in a manner that will be of practical use.

The writer doesn't know. Superficially at least, central determination of a single Order and Shipping Time and Safety Level, the current practice, seems unsuitable. There probably is no substitute for down-to-earth hard drudgery in investigating various supply systems by persons thoroughly familiar in the technical sense with the material involved. An approach worth exploring is one that would authorize activities to compute their own Order and Shipping Times on the basis of own experience for different items and to be authorized Safety Levels which would give about the same probability of satisfaction to all. It might be practicable to have only two different Order and Shipping Times in any one activity; e.g., 10 days for small items sent parcel post and 30 days for bulky items sent by railway freight. Then by centrally prescribing a scale of Safety Levels varying with say three or four Order and Shipping Times, achievement of the same probability of depletion together with lowest total inventory might be attainable. For example, with a 3 months Operating Level and providing protection equal to 1-1/2 times average use, Safety Levels would be as follows:

	Months Order & Shipping Time			
	.25	.5	1.0	1.5
Months Safety Level	.125	.25	.5	.75
Months Inventory at Order Point	.375	.75	1.5	2.25

Of course, many other considerations are involved. Some items have such erratic and unpredictable use that a much higher protection than 1-1/2 times average usage is necessary. Other items have such small usage that unit containers exceed not only conventional Safety Levels but

even Operating Levels. A few items are so important that depletion must not be allowed to happen regardless of cost. Such considerations may, upon investigation, negate any possibility of taking advantage of such approaches or may limit applicability to very large stations only. Nevertheless, the drudgery of a thorough study appears worth-while in view of the possibilities.

SUMMARY AND CONCLUSIONS

A superficial study of some factors directly and indirectly related to stock levels and of some collateral associated logistic considerations has been presented. Suggestions as to their practical exploitation were advanced. It was observed that the arbitrary imposition of logistic policies; such as too small inventories, by fiscal and budgetary decisions may inhibit true economy and may impair efficiency.

* * *

THE ROMMEL PAPERS – A COMMENTARY[1]

Rear Admiral Henry E. Eccles
U. S. Navy (Retired)

This article is a commentary on The Rommel Papers, edited by B. H. Liddell Hart with the assistance of Lucie-Maria Rommel, Manfred Rommel, and General Fritz Bayerlein; translated by Paul Findlay. New York: Harcourt, Brace, and Company, 1953, 524 pages.

No student of war can afford to neglect the remarkable story of World War II as seen by "one of the Great Captains," for in The Rommel Papers are contained, in extraordinarily forceful and cogent form, some of the most important lessons in command, strategy, logistics, tactics, sea power and mobile warfare yet published.

From the excellent analytical introduction, which B. H. Liddell Hart concludes with the appraisal that Rommel is truly one of the "Great Captains," to Rommel's final, tragic, sentence, "The sky over Germany has grown very dark," this book is packed with illustrations of both the understanding use and the flagrant violation of sound principles for the conduct of war.

The fact that this fascinating narrative moves so compellingly is a tribute, not only to the genius of Rommel, but, also, to the understanding and skill of the editor and of the translator. The former, Capt. B. H. Liddell Hart, has long been known, particularly in Europe, for his pioneering thinking in the use of armored forces. The true depth of his understanding has only recently been recognized in this country. Certainly the masterly and restrained notations which he intersperses throughout the narrative add greatly to the value of the book. Furthermore, they form a refreshing contrast to the voluminous footnotes that so frequently distract the reader of military commentaries. This proves that editorial scholarship is not necessarily dull. These notations in many instances supply information as to the British situation that Rommel lacked, and in other instances, shrewdly point up lessons in the art of war that are most pertinent.

The translation by Paul Findlay is so skillful that the American reader catches the full feeling of the man and the campaigns, without being aware that these clear accounts and comments were originally written in a foreign language and grammar.

The excellence of the maps, some of them the actual sketches by Rommel himself, adds greatly to the value of the book, for their simplicity and clarity keep the reader ever mindful of the relative location of the physical objectives without drowning him in detail.

Towering over every other aspect of the book is the extraordinary character of Rommel the leader, who in spite of his deep interest and understanding of the problems and techniques of tactics, always saw his actions against the background of strategy, and, even when he knew his cause was hopeless, never lost his fighting spirit. He was a man who despised the arm chair strategist and yet was himself a deep student who contributed greatly to the understanding of war by his penetrating analyses written at every opportunity during lulls in his fighting.

---

[1]This article was prepared under sponsorship of the office of Naval Research under contract N7onr-41904 with The George Washington University.

The first part of the book deals with the campaign in France in 1940 and really serves as an introduction to the main part, the African Campaign, where the tides of battle ebbed and flowed from 1940 to 1943. The last part of the story is that of the invasion of Normandy and in October 1944 the final, tragic, murderous, suicide of a great and loyal soldier.

Throughout this tale, the major lessons of the art of war ring out like the sound of great bells tolling over the confused noise of traffic in a large city, ignored by the vast multitude unconcerned in their pursuit of momentary ends, but, clear and deeply significant to those whose minds are tuned to their depth and meaning.

To the student of logistics, it is interesting to note that while a great part of the book is devoted to the discussion of logistic problems and deficiencies and their tremendous influence on both tactics and strategy, the word "logistics" is not once used. The index mentions the word twice: once referring to page 102, again to page 264, both references being merely comments on problems of transportation. Yet, logistics, in the sense the word is now being used in the United States, was a transcending factor in every campaign.

This inadequacy of concept on the part of both the general and the editor is not due to any lack of understanding as to the immediate effect of inadequate logistic planning and support upon the field commander, but rather it arises from a lack of understanding on their part as to the fundamental causes of inadequate logistic planning and support.

This is a defect shared by many because it is only recently that any serious studies have been undertaken in the field of logistics research, and it is only through the vigorous prosecution of these studies that the nature of these cause and effect relationships can be determined.

That General Rommel understood the general outline of the problem is indicated by two passages, best read in reverse sequence.

Writing about the evacuation of Cyrenaica in November of 1942 he says, on page 352:

"Marshal Cavallero had been in Libya since the 12th November, but had not, of course, thought it necessary to come and confer with me, although I had several times asked him to. He sent me instead, via Ritter von Pohl (German Air Attache), an order on behalf of the Duce to Stand fast in Cyrenaica for at least another week. The same order also required us to hold on to the Mersa el Brega position at all costs, as the fate of the Axis forces in Africa depended on it. It would have been very much better if Marshal Cavallero had applied this zeal he was now asking from us to the task of supplying the Panzer Army before Alamein, or, indeed, if he had now begun to do so. We had always done our share, and the courage of my troops had overcome difficulties far worse than anything the supply staffs had had to meet. It was time that our higher authorities demanded the same expenditure of energy from themselves as we had made our norm. Marshal Cavallero belonged to the type of the intellectually fairly well-qualified, but weak-willed office-chair soldier. The organization of supplies, the command of men, anything in any way constructive requires more than intellect; it requires energy and drive and an unrelenting will to serve the cause, regardless of one's personal interests. Academic soldiers mostly look on war as a pure intellectual problem and demand energy and drive only from those whom they somewhat contemptuously refer to as "Troupier," (A term, usually of scorn, used to denote the field commander with few or no staff or academic qualifications.) but from themselves, never. They rest content with their professional qualifications, which are attested for them by others of the same ilk, and regard themselves as the source of all good and the "Troupier" as the source of all evil. It is time that a clean sweep was made of this mentality, both in Germany and Italy."

Earlier in the book on page 92, in writing about the defeat of Graziani in November 1940, Rommel says:

"If quartermasters and civilian officials are left to take their own time over the organisation of supplies, everything is bound to be very slow. Quartermasters often tend to work by theory and base all their calculations on precedent, being satisfied if their performance comes up to the standard which this sets. This can lead to frightful disasters when there is a man on the other side who carries out his plans with greater drive and thus greater speed. In this situation the commander must be ruthless in his demands for an all-out effort. If there is anyone in a key position who appears to be expending less than the energy that could properly be demanded of him, or who has no natural sense for practical problems of organisation, then that man must be ruthlessly removed. A commander must accustom his staff to a high tempo from the outset and continually keep them up to it. If he once allows himself to be satisfied with norms, or anything less than an all-out effort, he gives up the race from the starting-post and will sooner or later be taught a bitter lesson by his faster-moving enemy and be forced to jettison all his fixed ideas."

and later on pages 96 and 97, he says:

"When a commander has won a decisive victory- and Wavell's victory over the Italians was devastating - it is generally wrong for him to be satisfied with too narrow a strategic aim. For that is the time to exploit success. It is during the pursuit, when the beaten enemy is still dispirited and disorganised, that most prisoners are made and most booty captured. Troops who on one day are flying in a wild panic to the rear, may, unless they are continually harried by the pursuer, very soon stand in battle again, freshly organised as fully effective fighting men.

The reason for giving up the pursuit is almost always the quartermaster's growing difficulty in spanning the lengthened supply routes with his available transport. As the commander usually pays great attention to his quartermaster and allows the latter's estimate of the supply possibilities to determine his strategic plan, it has become the habit for quartermaster staffs to complain at every difficulty, instead of getting on with the job and using their powers of improvisation, which indeed are frequently nil. But generally the commander meekly accepts the situation and shapes his actions accordingly.

When, after a great victory which has brought the destruction of the enemy, the pursuit is abandoned on the quartermaster's advice, history almost invariably finds the decision to be wrong and points to the tremendous chances which have been missed. In face of such a judgment there are, of course, always academic soldiers quick to produce statistics and precedents by people of little importance to prove it wrong. But events judge otherwise, for it has frequently happened in the past that a general of high intellectual powers has been defeated by a less intelligent but stronger willed adversary.

The best thing is for the commander himself to have a clear picture of the real potentialities of his supply organisation and to base all his demands on his own estimate. This will force the supply staffs to develop their initiative, and though they may grumble, they will as a result produce many times what they would have done left to themselves."

In other words, Rommel in these passages both raises and almost unknowingly answers several of the gravest questions in military command and organization. These are in essence the questions that are at the bottom of many of the most difficult reorganizational maneuvers that have taken place in Washington since 1946.

How do we achieve the same understanding of strategy and sense of urgency in our logistical planning and administration as we find in our best field commanders? How do we get the basic integration of strategic and logistic planning that is essential at all levels of command? For the failure to achieve the answers to these questions is one of the four basic reasons why when Germany had decisive victory in her grasp it slipped away. The other three reasons (even more basic than this major logistical fault and also brought out clearly throughout the Rommel Papers) are: the failure of the German Feuhrer and OKW to select the correct strategical objectives, their failure to gain command of the air, and, their failure to understand the nature and employment of sea power.

The Rommel Papers present an excellent opportunity for some research student to examine the workings of the "Fourth Logistic Service" as exemplified by the traditional Italian Intendente against the light of General Rommel's experience with it in Africa and its workings in Greece.

No command lesson in history is clearer than the constant reiteration by Rommel of the disastrous results which can follow when the combat commander does not control his own logistics. Pages, 138, 191, 192 and 250 not only illustrate this principle, but, they also illustrate the valuable subsidiary logistics principle that the theater transportation system must be responsive to the needs of the combat commander.

> "A further trial for us was that the Italians were still carrying the bulk of our supplies to Tripoli and making very little use of Benghazi. Tripoli was over 1,000 miles from the front. Bearing in mind that 1,500 tons of supplies, including water and rations, had to be carried up to the front every day, even for normal activity, it is easy to understand that our transport could not cope indefinitely with a route of that length. With no authority over the people responsible for shipping in the Mediterranean, however, it was very difficult for us to do anything about it.
>
> "It was obvious that the High Command's opinion had not changed from that which they had expressed in 1941, namely, that Africa was a "lost cause" and that any large-scale investment of material and troops in that theatre would pay no dividends. A sadly short-sighted and misguided view! For, in fact, the supply difficulties which they were so anxious to describe as "Insuperable" were far from being so. All that was wanted was a real personality in Rome, someone with the authority and drive to tackle and clear away the problems involved. No doubt it would have led to friction in certain Italian circles, but this could have been overcome by an authority unencumbered with other political functions. Our Government's weak policy towards Italy seriously prejudiced the German-Italian cause in North Africa.
>
> "Unfortunately, the refit of our formations made very slow progress, due to the fact that for some unaccountable reason the few ships engaged on the Africa run were still not arriving at Tobruk or Mersa Matruh but at Benghazi or Tripoli. This meant that all supplies had to be carried by transport columns or our few coastal vessels, over a distance of either 750 or 1,400 miles. This, of course, was more than we could manage."

The constant cry for petrol, the discussions of how its lack, time and again, defeated his opportunities, and his anger at the failure to supply him adequately, cannot be laid to unreasonableness on Rommel's part, nor to lack of ability of higher authority to supply it. It was almost wholly a matter of the failure of higher authority to grasp the realities of the situation. This is conclusively proven by the fact that once the battle for Africa was hopelessly lost, the necessary petrol and equipment were found and sent. But, the time had passed when it could have

been unloaded and distributed, and by that time the Anglo-American strength had been so built up that nothing could then avail.

Perhaps the most significant lesson to be derived from this is: If, logistical considerations are to be properly evaluated in strategic planning and, if, a flexible and responsive logistical system is to be created, then it is essential that the logistics of a theater of war be seen as an integrated whole rather than as casually connected but separate parts.

There can be no satisfactory substitute for a thorough study of this book; for, in many parts, there are messages that can be fully grasped only by re-reading them in the context of subsequent chapters.

However, a few points seem to stand out as particularly worthy of note. Of great importance is the story of the relations between, and the relative effectiveness of, German and Italian troops. Rommel discusses this with scrupulous fairness and great insight, which is a refreshing contrast to the superficial contempt of some contemporary writers.

On page 97 there is a splendid discussion of the effects of Graziani's defeat on the confidence and morale of Italian troops, and, on pages 243, 244, 261, and, 262, there are analyses of the basic causes of the defeats Italian forces so frequently suffered. Rommel's exasperated contempt for the short sightedness and interference of the Commando Supremo is matched by his generous appreciation for the courage and ability of many Italian officers and units.

Rommel's comments on the art of command thinking are frequent and to the point. An example is:

> "However praiseworthy it may be to uphold tradition in the field of soldierly ethics, it is to be resisted in the field of military command. For to-day it is not only the business of commanders to think up new techniques which will destroy the value of the old: the potentialities of warfare are themselves being continually changed by technical advance. Thus the modern army commander must free himself from routine methods and show a comprehensive grasp of technical matters, for he must be in a position continually to adapt his ideas of warfare to the facts and possibilities of the moment. If circumstances require it, he must be able to turn the whole structure of his thinking inside out.
>
> I think that my adversary, General Ritchie, like so many generals of the old school, had not entirely grasped the consequences which followed from the fully motorised conduct of operations and the open nature of the desert battlefield. In spite of the good detailed preparation of his plans, they were bound to go wrong, for they were, in essence, a compromise."

Throughout his comments he is generous and warm in his praise of the virtues of enemy commanders, although, quick to point out their errors and weaknesses.

The inclusion of a clear description of the composition and relationship of the various elements of the command structure such as the OKW, the Commando Supremo, and the Governor of Libya, would add greatly to the value of the book to the military student. This is a matter which can easily be, and should be, corrected in the next edition of this book.

Summing up, "The Rommel Papers" is a great book, one of the great commentaries of military literature. To each student it provides solid nourishment and challenge.

To all military readers it shows the interweaving of the various elements of war, either by their actions and interactions or by their neglect: grand strategy, logistics, tactics, organization and command relations, and above all, leadership, all are included.

To the strategist there is the challenge of analyzing the choice of objectives and the use of sea power.

To the logistician, there is the challenge of how best to provide a system and administration of logistic planning and support worthy of the genius of a great fighter.

To the tactician there is the challenge of meeting the need for flexibility and imagination in the employment of old and new weapons.

To the government leader there is the challenge of providing a national security organization and a system of military alliances that are adequate to the national requirements.

To the historian there is the fascinating example of the living and the dead, the teacher and the pupil, the scholar and the fighting general, Liddell Hart and Rommel, collaborating in producing a work of historical art.

To the philosopher there is the contemplation of the deep tragedy of so great and generous a man and general devoting his life to such a contemptible cause and to such a depraved leader.

* * *

A NUMERICAL METHOD TO DETERMINE OPTIMUM STRATEGY

John von Neumann
Institute for Advanced Study

The paper that follows describes a numerical method to solve "2-person game" or "optimum linear programming" problems, in such a manner that both the maximum duration of the calculation and the maximum size of numbers that might occur in the course of it can be guaranteed in advance. Some remarks on the comparison of this method with G. Dantzig's "Simplex Method" are given at the end of the paper.

1. It is desired to find a numerical procedure to solve the "Minimax" or "two person game" or "optimum linear program" problem for the rectangular "payoff" matrix

(1) $$[[a_{ij}]] \quad (i = 1, \ldots, m;\ j = 1, \ldots, n).$$

This means to find two "mixed strategies," i.e. vectors

(2) $$[x_i] \quad (i = 1, \ldots, m)$$
$$\text{all } x_i \geqq 0,\ \Sigma_i x_i = 1,$$

and

(3) $$[y_j] \quad (u = 1, \ldots, n)$$
$$y_j \geqq 0,\ \Sigma_j y_j = 1,$$

and a "value" k, such that

(4) $$\text{all } \Sigma_i a_{ij} x_i \geqq k,$$
$$\text{all } \Sigma_j a_{ij} y_j \leqq k.$$

Consider any pair of vectors $[x_i]$, $[y_j]$ fulfilling (2), (3), and any k. Form the vectors $[U_i]$, $[V_j]$ with

(5) $$U_i = \Sigma_j a_{ij} y_j - k,$$
$$V_j = k - \Sigma_i a_{ij} x_i,$$

then the vectors $[u_i]$, $[v_j]$ with

(6) $$u_i = \text{Max } (0, U_i),$$
$$v_j = \text{Max } (0, V_j),$$

and finally the scalar ϕ with

(7) $$\phi = \Sigma_i u_i^2 + \Sigma_j v_j^2.$$

Then (4) means

(8) $$\text{all } U_i \leqq 0,\ \text{all } V_j \leqq 0,$$

i.e.

(9) $$\text{all } u_i = 0,\ \text{all } v_j = 0,$$

i.e.

(10) $$\phi = 0.$$

Hence solving (4) (with (2), (3)) is equivalent to reducing ϕ to its minimum value (under these assumptions), which is 0.

2. Consider a system $[\bar{x}_i]$, $[\bar{y}_j]$, $\bar{k}$, also satisfying (2), (3). With a suitable θ,

(11) $$0 \leqq \theta \leqq 1$$

we will form a third system $[x_i']$, $[y_i']$, k', also satisfying (2), (3), according to

(12) $$\begin{aligned} x_i' &= (1-\theta)\, x_i + \theta\, \bar{x}_i, \\ y_j' &= (1-\theta)\, y_j + \theta\, \bar{y}_j, \\ k' &= (1-\theta)\, k + \theta\, \bar{k}. \end{aligned}$$

Form $[\bar{U}_i]$, $[\bar{V}_j]$ from $[\bar{x}_i]$, $[\bar{y}_j]$, $\bar{k}$ according to (5), and $[U_i']$, $[V_j']$, $[u_i']$, $[v_j']$, ϕ' from $[x_i']$, $[y_j']$, k' according to (5)-(7). We wish to choose θ so as to minimize ϕ'. In the course of this process we will also make a suitable selection of $[\bar{x}_i]$, $[\bar{y}_j]$, $\bar{k}$.

3. In view of (5), (6)

$$U_i' \leqq 0 \text{ implies } u_i' = 0,$$

$$U_i' > 0 \text{ implies } u_i' = U_i', \text{ hence}$$

$$0 < u_i' = U_i' = (1-\theta)\, U_i + \theta\, \bar{U}_i \leqq (1-\theta)\, u_i + \theta\, \bar{U}_i.$$

Hence in any event

$$u_i'^2 \leqq ((1-\theta)\, u_i + \theta\, \bar{U}_i)^2.$$

Similarly

$$v_j'^2 \leqq ((1-\theta)\, v_j + \theta\, \bar{V}_j)^2,$$

and therefore by summation

$$\phi' \leqq \Sigma_i\, ((1-\theta)\, u_i + \theta\, \bar{U}_i)^2 + \Sigma_j\, ((1-\theta)\, v_j + \theta\, \bar{V}_j)^2 = (1-\theta)^2\, (\Sigma_i\, u_i^2 + \Sigma_j\, v_j^2)$$
$$+ 2\,\theta\, ((1-\theta)\, (\Sigma_i\, u_i\, \bar{U}_i + \Sigma_j\, v_j\, \bar{V}_j) + \theta^2\, (\Sigma_i\, \bar{U}_i^2 + \Sigma_j\, \bar{V}_j^2),$$

i.e.

(13) $$\phi' \leqq \phi\, (1-\theta)^2 + 2B\,\theta\, (1-\theta) + C\theta^2,$$

where

$$B = \Sigma_i\, u_i\, \bar{U}_i + \Sigma_j\, v_j\, \bar{V}_j,$$

$$C = \Sigma_i\, \bar{U}_i^2 + \Sigma_j\, \bar{V}_j^2.$$

More explicitly, by (5),

(14) $$B = \Sigma_{ij}\, a_{ij}\, (u_i\, \bar{y}_j - v_j\, \bar{x}_i) - (\Sigma_i\, u_i - \Sigma_j\, v_j)\, \bar{k}.$$

Now put

(15) $$\bar{x}_i = \frac{u_i}{\Sigma_i\, u_i}, \qquad \bar{y}_j = \frac{v_j}{\Sigma_j\, v_j}.$$

These $[\bar{x}_i]$, $[\bar{y}_j]$ fulfill (2), (3), of course. If $\Sigma_i\, u_i = 0$, i.e. if all $u_i = 0$, choose any $[\bar{x}_i]$ fulfilling (2), (3). If $\Sigma_j\, v_j = 0$, i.e. if all $v_j = 0$, choose any $[\bar{y}_j]$ fulfilling (2), (3). Then in any event

$$u_i = \bar{x}_i\, \Sigma_i\, u_i, \qquad v_j = \bar{y}_j\, \Sigma_j\, v_j.$$

Hence (14) gives

(16) $$B = (\Sigma_{ij}\, a_{ij}\, \bar{x}_i\, \bar{y}_j - \bar{k})\, (\Sigma_i\, u_i - \Sigma_j\, v_j).$$

Next put

(17) $$\bar{k} = \Sigma_{ij}\, a_{ij}\, \bar{x}_i\, \bar{y}_j.$$

Then (16) implies

(18) $$B = 0,$$

hence (13) becomes

(19) $$\phi' \leqq \phi (1 - \theta)^2 + C\theta^2.$$

The right hand side of (19) is minimized when

$$-2\phi (1 - \theta) + 2C\theta = 0,$$

i.e. for

(20) $$\theta = \frac{\phi}{\phi + C}.$$

Note, that $\phi \geqq 0$, $C \geqq 0$ by (7), (13), and we assume $\phi \neq 0$ — since otherwise the problem (of making $\phi = 0$) would already be solved. Hence $\phi > 0$, therefore θ can be chosen according to (20), and it satisfies (11). We are, therefore, choosing θ according to (20).

Now (19), (20) give

(21) $$\phi' \leqq \frac{\phi C}{\phi + C},$$

i.e.

(22) $$\frac{1}{\phi'} \geqq \frac{1}{\phi} + \frac{1}{C}.$$

4. Put

(23) $$\mu = \mathrm{Min}_{ij} (a_{ij}), \quad \nu = \mathrm{Max}_{ij} (a_{ij}).$$

Then (2), (3) give

$$\mu \leqq \Sigma_i a_{ij} \bar{x}_i, \quad \Sigma_j a_{ij} \bar{y}_j \leqq \nu,$$

and (2), (3), (17) give

$$\mu \leqq \Sigma_{ij} a_{ij} \bar{x}_i \bar{y}_j = \bar{k} \leqq \nu.$$

Hence (5) gives

$$-(\nu - \mu) = \mu - \nu \leqq U_i, \quad V_j \leqq \nu - \mu,$$

i.e.

$$U_i^2, V_j^2 \leqq (\nu - \mu)^2,$$

and so, by (13),

(24) $$C \leqq (m + n) (\nu - \mu)^2.$$

Consequently (22) becomes

(25) $$\frac{1}{\phi'} \geqq \frac{1}{\phi} + \frac{1}{(m + n) (\nu - \mu)^2}.$$

5. Now let us assume, that the process of 2. - 4., which lead from the system $[x_i]$, $[y_j]$, k to the system $[x_i']$, $[y_j']$, k', is being applied sequentially, yielding successively the systems

(26) $$[x_i'], [y_j'], k'; [x_i''], [y_j''], k''; [x_i'''], [y_j'''], k'''; \text{ etc., etc.}$$

Denote the h-th system (h = 1, 2, ...) in the sequence of (26) by $[x_i^{(h)}]$, $[y_j^{(h)}]$, $k^{(h)}$. Form its $[U_i^{(h)}]$, $[V_j^{(h)}]$, $[u_i^{(h)}]$, $[v_j^{(h)}]$, $\phi^{(h)}$ according to (5)-(7). Then h-fold application of (25) gives

$$\frac{1}{\phi^{(h)}} \geqq \frac{h}{(m + n) (\nu - \mu)^2},$$

i.e.

$$\phi^{(h)} \leqq \frac{(m+n)(\nu-\mu)^2}{h}. \tag{27}$$

In view of (7) this can also be written

$$\Sigma_i (u_i^{(h)})^2 + \Sigma_j (v_j^{(h)})^2 \leqq \frac{(m+n)(\nu-\mu)^2}{h},$$

i.e.

$$\sqrt{\frac{\Sigma_i (u_i^{(h)})^2 + \Sigma_j (v_j^{(h)})^2}{m+n}} \leqq \frac{\nu-\mu}{\sqrt{h}} \tag{28}$$

Note, that the left hand side of (28) is the quadratic mean of all the $u_i^{(h)}$ and all the $v_j^{(h)}$ — this, then, is by (28) at most the $\sqrt{h}$-th part of $\nu - \mu$, i.e. of the total variation of the a_{ij}. Note, that $\nu - \mu$, the difference between the largest possible payoff and the smallest possible payoff, represents the total risk involved in this game. Note, also, that the $[u_i^{(h)}]$, $[v_j^{(h)}]$ express, how far away the $[x_i^{(h)}]$, $[y_i^{(h)}]$, $k^{(h)}$ still are from a solution.

Regarding the latter, consider (4) and (9), and the statement of their equivalence. Note also (dropping the $^{(h)}$), that by (5), (6)

$$\begin{aligned} &\mathrm{Max}_i (\Sigma_j a_{ij} y_j) = k + \mathrm{Max}_i (U_i) = k + \mathrm{Max}_i (u_i), \\ &\mathrm{Min}_j (\Sigma_i a_{ij} x_i) = k - \mathrm{Max}_j (V_j) = k - \mathrm{Max}_j (v_j). \end{aligned} \tag{29}$$

Since the true "value" of the game is necessarily $\leqq \mathrm{Max}_i (\Sigma_j a_{ij} y_j)$ and $\geqq \mathrm{Min}_j (\Sigma_i a_{ij} x_i)$, therefore the approximate value k cannot err by more than $\mathrm{Max}_i (U_i) = \mathrm{Max}_i (u_i)$ down or $\mathrm{Max}_j (V_j) = \mathrm{Max}_j (v_j)$ up.

$h \to \infty$ furnishes a convergent process, which gives a solution $[x_i]$, $[y_j]$, k of (4) as a limit of the sequence (or rather of a suitable subsequence of) $[x_i^{(h)}]$, $[y_j^{(h)}]$, $k^{(h)}$. We will not go into this here.

6. Let us summarize the computational process, which leads from $[x_i]$, $[y_j]$, k to $[x_i']$, $[y_j']$, k', according to 2. - 4.

It is advisable to carry $[U_i]$, $[V_j]$, too, along in this inductive step. Indeed, if we do not do this, then forming $[U_i]$, $[V_j]$ and $[\overline{U}_i]$, $[\overline{V}_j]$ requires, owing to (5), four applications of the matrix $[[a_{ij}]]$ to suitable vectors (namely to $[x_i]$, $[y_j]$ and $[\overline{x}_i]$, $[\overline{y}_j]$). If, on the other hand, $[U_i]$, $[V_j]$ are carried along in the inductive step, then only two such applications will be required. Indeed, forming $[\overline{U}_i]$, $[\overline{V}_j]$ requires, owing to (5), two applications of $[[a_{ij}]]$ to suitable vectors (namely to $[\overline{x}_i]$, $[\overline{y}_j]$). $[U_i']$, $[V_j']$, on the other hand, can be obtained differently. Indeed, (12) gives

$$\begin{aligned} U_i' &= (1-\theta) U_i + \theta \overline{U}_i, \\ V_j' &= (1-\theta) V_j + \theta \overline{V}_j. \end{aligned} \tag{30}$$

In view of this, the inductive step, based on (6), (15), (17), (5), (7), (13), (20), (12), (30), is as follows.

We start with

$$[x_i], [y_j], k, [U_i], [V_j] \tag{31}$$

given. ($[U_i]$, $[V_j]$ are actually determined by $[x_i]$, $[y_j]$, k according to (5), but this is irrelevant for the moment.) The following calculations must be carried out.

$$u_i = \mathrm{Max}(0, U_i), \tag{32.1}$$

$$v_j = \mathrm{Max}(0, V_j), \tag{32.2}$$

$$s = \Sigma_i u_i, \tag{32.3}$$

$$t = \Sigma_j v_j, \tag{32.4}$$

$\bar{x}_i = u_i/s$ unless $s = 0$, then, say, $\bar{x}_i = x_i$, (32.5)

$\bar{y}_j = v_j/t$ unless $t = 0$, then, say, $\bar{y}_j = y_j$, (32.6)

$$\bar{S}_i = \Sigma_j a_{ij} \bar{y}_j, \tag{32.7}$$

$$\bar{T}_j = \Sigma_i a_{ij} \bar{x}_i, \tag{32.8}$$

$$\bar{k} = \Sigma_i \bar{S}_i \bar{x}_i = \Sigma_j \bar{T}_j \bar{y}_j, \tag{32.9}$$

$$\bar{U}_i = \bar{S}_i - \bar{k}, \tag{32.10}$$

$$\bar{V}_j = \bar{k} - \bar{T}_j, \tag{32.11}$$

$$\phi = \Sigma_i u_i^2 + \Sigma_j v_j^2, \tag{32.12}$$

$$C = \Sigma_i \bar{U}_i^2 + \Sigma_j \bar{V}_j^2, \tag{32.13}$$

$$\theta = \phi / (\phi + C), \tag{32.14}$$

$$x_i' = x_i - \theta (x_i - \bar{x}_i), \tag{32.15}$$

$$y_j' = y_j - \theta (y_j - \bar{y}_j), \tag{32.16}$$

$$k' = k - \theta (k - \bar{k}), \tag{32.17}$$

$$U_i' = U_i - \theta (U_i - \bar{U}_i), \tag{32.18}$$

$$V_j' = V_j - \theta (V_j - \bar{V}_j). \tag{32.19}$$

(32.15)-(32.19) furnish

$$[x_i'], [y_j'], k', [U_i'], [V_j'], \tag{33}$$

and thereby complete the inductive step.

The operation counts in these formulae (32.1)-(32.19) are as follows:

	Max (0, ...)	Additions and Subtractions	Multiplications	Divisions
(32.1) to (32.19), other than (32.7) and (32.8)	m + n	8(m + n) + (m or n) - 4	4 (m + n) + (m or n) + 1	m + n + 1
(32.7)	--	mn - m	mn	--
(32.8)	--	mn - n	mn	--
Total	m + n	2 mn + 7(m + n) + (m or n) - 4	2 mn + 4(m + n) + (m or n) + 1	m + n + 1

(34)

This is, therefore, the operation time of about

$$N = 2\,mn + 5\,(m + n) + (m \text{ or } n) + 2 \tag{35}$$

multiplications and their normal complement. Under the conditions of the "701" this is about N msec, under those of the "IAS" machine about 2N msec. The only large memory requirement is represented by $[[a_{ij}]]$, this may be stored on a drum, and at 1/2 precision ("701" — 17 bits, i.e. $2^{-17} = 7.6 \times 10^{-6}$) or 1/4 precision ("IAS" — 9 bits, i.e. $2^{-9} = 2 \times 10^{-3}$). The problem can be organized in such a way, that the drum does not cause a relevant slowing down. Actually (32.7), (32.8), which alone require the extra memory capacity of the drum, are best treated by subroutines. These subroutines are also otherwise useful, e.g. to check (5.)

Concerning the sizes of the numbers that may occur in (33)-(35), we note this. Assume

(36) $$-1/2 \leqq \mu \leqq \nu \leqq 1/2,$$

(cf. (23)), i.e.

(37) $$-1/2 \leqq a_{ij} \leqq 1/2.$$

Then (2), (3) for $[x_i]$, $[y_j]$ imply

(38) $$0 \leqq x_i,\ y_j \leqq 1.$$

It is nonsensical to violate

(39) $$\mu \leqq k \leqq \nu$$

(k is an approximant of the "value"!), hence (36) gives

(40) $$-1/2 \leqq k \leqq 1/2.$$

(5), with (2), (3) and (37), guarantees

$$-1/2 \leqq \Sigma_i a_{ij} x_i,\quad \Sigma_j a_{ij} y_j \leqq 1/2,$$

hence, using (40), too,

(41) $$-1 \leqq U_i,\quad V_j \leqq 1.$$

Now clearly (using (32.1)-(32.19) and the above)

(42.1) $$0 \leqq u_i,\quad v_j \leqq 1,$$

(42.2) $$0 \leqq s \leqq m,\quad 0 \leqq t \leqq n,$$

(42.3) $$[\overline{x}_i],\ [\overline{y}_j] \text{ fulfill (2), (3),}$$

(42.4) $$0 \leqq \overline{x}_i,\quad \overline{y}_j \leqq 1,$$

(42.5) $$-1/2 \leqq \mu \leqq \overline{S}_i,\quad \overline{T}_j \leqq \nu \leqq 1/2,$$

(42.6) $$-1/2 \leqq \mu \leqq \overline{k} \leqq \nu \leqq 1/2,$$

(42.7) $$-1 \leqq \overline{U}_i,\quad \overline{V}_j \leqq 1,$$

(42.8) $$0 \leqq \phi \leqq m + n,$$

(42.9) $$0 \leqq C \leqq m + n,$$

(42.10) $$0 \leqq \theta \leqq 1,$$

(42.11) $$[x_i'],\ [y_j'] \text{ fulfill (2), (3),}$$

(42.12) $$0 \leqq x_i',\quad y_j' \leqq 1,$$

(42.13) $$-1/2 \leqq \mu \leqq k' \leqq \nu \leqq 1/2,$$

(42.14) $$-1 \leqq U_i',\quad V_j' \leqq 1.$$

Thus (42.13), (42.14) reconfirm (39)-(41) for k', $[U_i']$, $[V_j']$, and, of course, (42.11) reconfirms (2), (3) for $[x_i']$, $[y_j']$. Hence all of these assumptions remain valid throughout the induction.

Note, that (42.1)-(42.14) imply, that only s, t, ϕ, C need be scaled, and even those only by fixed factors, which are known ahead of time.

7. To conclude, we should make some stipulation about the first system (31), i.e. the one that starts the entire inductive process.

$[x_i]$, $[y_j]$ should be chosen in some plausible way, of course satisfying (2), (3). If no more plausible cue is available, the "equipartitioned" choice

(43) $$\text{all } x_i = \frac{1}{m},\ \text{all } y_j = \frac{1}{n}$$

suggests itself. Next, in order to get to the $[U_i]$, $[V_j]$ according to (5),

$$S_i = \Sigma_j a_{ij} y_j, \tag{44.1}$$

$$T_j = \Sigma_i a_{ij} x_i, \tag{44.2}$$

must be formed. Now for k some choice

$$\mathrm{Min}_j (T_j) \leqq k \leqq \mathrm{Max}_i (S_i) \tag{44.3}$$

is appropriate. Now (5) requires

$$U_i = S_i - k, \tag{44.4}$$

$$V_j = k - T_j. \tag{44.5}$$

Note, that (44.1), (44.2) imply, in view of (2), (3),

$$-1/2 \leqq \mu \leqq S_i, \quad T_j \leqq \nu \leqq 1/2. \tag{45.1}$$

From this, by (44.3),

$$-1/2 \leqq \mu \leqq k \leqq \nu \leqq 1/2. \tag{45.2}$$

Next (44.4), (44.5) give

$$-1 \leqq U_i, \quad V_j \leqq 1. \tag{45.3}$$

(45.2), (45.3) reconfirm (39)-(41). (45.1)-(45.3) show that no scaling problems arise here.

In evaluating the method it ought to be compared with G. Dantzig's "Simplex Method." In the latter method the a priori guarantees for length of calculation and size of numbers are considerably less favorable than ours, but the available practical experience with the simplex method indicates that its actual performance — at least under the conditions under which it has been so far tested — is much better than the limits that can be guaranteed. Limited comparisons of our method with the "simplex method" again indicate that the latter converges faster, but there is reason to believe that our method can be accelerated by various tricks which amount to smoothing the iterative recursive sequence which is involved, and making this recursion dependent on several predecessors. The description of the method is offered here as a first step in this direction, i.e. in order to illustrate the new method, and to furnish a basis for the possible improvements referred to above.

* * *

OPTIMAL WEAPON SYSTEMS[1]

J. Marschak
University of Chicago
and
M. R. Mickey
Iowa State College

This paper presents a procedure for selecting that weapon system from a class of weapon systems with a fixed cost, which would maximize the expectation of the military utility. Although certain simplifying assumptions are made which make the results not directly applicable to real military problems, the paper may help to clarify some basic ideas.

1. **THE PROBLEM:** Assume that all targets have the same strength of local defenses and are also identical in all other relevant respects. Let the damage, i.e., the number of targets destroyed, be x. It is a random variable. We shall describe its distribution approximately, by a probability density function f(x) which depends on the amount of the various weapons used, i.e., on the vector $y = (y_1, \cdots, y_n)$ where y_i is the quantity of weapon units of the i-th kind. Let the money cost of producing y be C(y) and let the military utility of the damage be u(x). We want to find y such that the expected military utility

$$Eu(x) \equiv \int_0^\infty u(x)f(x|y)dx \equiv U(y), \tag{1.1}$$

say, be a maximum, subject to the constraints

$$y_i \geq 0, \quad i = 1, \cdots, n \tag{1.2}$$

$$C(y) = K; \tag{1.3}$$

where K is the total amount of money available (the "budget"). The optimal values of the vector y and of the expected utility U will be functions of K. We want to find these functions. The problem will be solved under certain assumptions. For a more general formulation see Sec. 11.

2. MILITARY UTILITY

The symbols in (1.1) can be interpreted in a more general fashion: x being an element of a set of "outcomes," and y being an element of a set of "decisions." It is asserted that for a rational decision-maker there must exist a function u(x) such that if y and y' are two

[1] Originally issued as RAND document S-8, November 1951 and presented at The George Washington University Logistics Conference in January 1952. The computations in Section 10 were made by Bernice Brown.

decisions, and if $U(y) > U(y')$, then y is chosen in preference to y'. This assertion follows from other, more innocently looking postulates on rational behavior,[2] of which the following is crucial: Let A, B and C be three objectives and suppose you are indifferent between A and B; then you must also be indifferent between two lottery tickets, the one offering A or C with certain odds, the other offering B or C with those same odds. [The objectives A, B, C may themselves be lottery tickets, i.e., represent probability distributions such as our f(x).]

Note that the rule of maximizing the expectation U of utility u(x) is, in general, not compatible with the more complicated rule offered occasionally, stating that one should take into account not only the expectation of utility but also, for example, its variance, in order to express the rational decision-maker's attitude to "risk." In the present approach, the attitude to risk is fully reflected in the decision-maker's utility function u(x), and hence in the expected utility U, as will be illustrated later in this section.

In the present paper, use will be made of the following military utility function of damage. Let α and β be, respectively, the lower and upper "thresholds of damage" in the following sense: if not more than α targets are destroyed, utility = 0; if β targets or more are destroyed, utility = 1; for intermediate levels of damage, interpolate linearly (see Fig. 1).

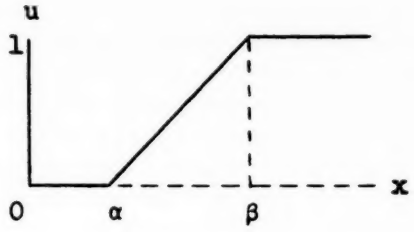

Figure 1

$$
\begin{aligned}
u(x) &= 0, && x < \alpha \\
&= \frac{x - \alpha}{\beta - \alpha}, && \alpha \leq x < \beta \qquad (2.1) \\
&= 1, && x \geq \beta .
\end{aligned}
$$

Instead of 0 and 1, we could have chosen any other two constants a, b ($a < b$). This would mean replacing the function u(x) by a linear increasing function of u(x). The expected value of this new function will be maximized by the same value of y as the expected value of the old function. However, 0 and 1 as the bounds for military utility suggest themselves because of a possible additional interpretation of military utility as the probability of victory. Whether the enemy does or does not surrender depends not only on the damage suffered but also on other factors. We may assume those other factors to have chance character so that, given the damage suffered (x), surrender has a certain probability u(x); and we want to maximize the expected value, U, of this probability.

---

[2] v. Neumann and Morgenstern; Marschak; Herstein and Milnor. (See list of references.)

By interpreting u(x) as the probability of victory we shall be able, in Section 3, to embed military utility in the more comprehensive concept of social utility. Those considerations help to understand and justify comparisons between the damage to the enemy and our expenditure in material resources and in lives. However, the rest of the paper is largely independent of those remarks.

No particular claim is attached to the function (2.1) except that this function seems to be ascertainable, albeit in a coarse fashion, with greater ease than some other suggested functions. A smooth function (differentiable for all $x > 0$) may be somewhat more realistic but its numerical characteristics are more elusive. In any case, we shall see that important properties of the suggested function are shared by all nonconstant and nondecreasing monotone functions, so that much of the results can be regarded simply as a convenient example for a more general case.

A more significant extension would be to make utility of damage decrease after a certain level of damage has been reached—if our concern is not only with the enemy's surrender but also with how to deal with the people of the conquered country. With this modification, u(x) could not be interpreted as the probability of victory. Also, the mathematical results, inasmuch as they are based on the assumption of u(x) being nondecreasing (see Sec. 4), would have to be revised.

In parts of this paper, the function u(x) defined in (2.1) will be further simplified by making $\alpha = \beta$ (as in Fig. 2).

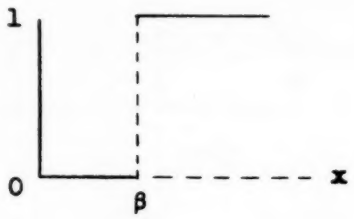

Figure 2

$$u(x) = 0, \quad x < \beta$$
$$= 1, \quad x \geq \beta. \tag{2.2}$$

The expected utility, U, is, in this case, equal to the probability that at least β targets are destroyed. Moreover, if u(x) is again interpreted as the probability of victory, (2.2) implies that victory is certain if at least β targets are destroyed, and is impossible otherwise.

The rule of maximizing expected utility Eu(x) (given the available money K, as a constraint) has been often abandoned in favor of maximizing the expectation of the number of targets destroyed, $Ex \equiv \int_0^\infty xf(x)dx$ (under the same constraint as above). Or, what is essentially the same thing, the cost of achieving a given mathematical expectation Ex was computed for each kind of weapon, or system of weapons, in order to choose the one that "costs least," in this particular sense. The recommendation used in the present paper can be stated as that of comparing the cost of achieving a given expectation of utility, Eu(x), with different weapons or weapon systems, in order to choose the one that costs least in this other sense. The present approach is more general, because maximizing Ex is equivalent to

maximizing Eu(x) in a special case—viz., when u(x) is linear. However, to assume u(x) linear is to neglect the fact that the "marginal utility" of the 200-th destroyed target is certainly different from that of the 10-th, even if all targets are physically identical. To maximize Ex may therefore involve a waste in resources. For example, if the utility function is approximately as in Fig. 1, then, for $x > \beta$ an additional dollar invested in the destruction of targets produces a smaller increase in the expected probability of victory than for $\alpha < x < \beta$. Hence, when larger amounts of money are available, it may become preferable to invest additional dollars elsewhere—yet the maximizer of Ex will always invest them in further destruction.

We shall denote by μ the expectation Ex of the number of targets destroyed, and by σ^2 its variance. If the utility function were linear, μ would be the only parameter of the distribution f(x) that would affect decisions. With a non-linear utility function other parameters become relevant. For example, if f(x) is (approximately) normal, the variance σ^2 becomes relevant along with μ.

Note that σ^2 is not necessarily an "undesirable thing" in the sense that, given μ, one should minimize σ^2. For example, using Fig. 2, compare two normal distributions of x, $f_1(x)$ and $f_2(x)$, with the same mean $\mu < \beta$, but different variances $\sigma_1^2 > \sigma_2^2$. Then the probability area assigned to values of x that exceed β and that therefore have the highest utility $u(\alpha) = 1$, rather than the lowest utility $u(\alpha) = 0$, is larger for f_1 than for f_2. If, on the other hand, $\mu > \beta$, then (for similar reasons) a distribution with lower variance is preferable. This point is worth making, as it is often stated that the variance of the number x of objects is a measure of "risk," and that risk is always undesirable. This statement would be correct if the utility function u(x) were everywhere strictly concave, i.e., if for any two points on the utility curve, the straight line connecting them were below the corresponding segment of the curve. (In the case of a differentiable utility function this means "decreasing marginal utility.")

3. REMARKS ON MILITARY AND SOCIAL UTILITY

Let us first consider the social utility w (for "welfare") as a function of only two "final goods": (1) victory V; (2) expenditure K, representing the decrease in the consumption of this and future generations. We shall show that the special problem of this paper, formulated in Section 1, is implied in the analogous but more general problem of maximizing the social utility w. Analogous with Section 1, replacing the objective x by the objectives V, K, and replacing the military utility u by the social utility w, we wish to maximize w with respect to the choice of strategy, i.e., of the variables under our (the government's) control. These variables are: (1) the expenditure K; (2) the general military policy, or, in a more restricted problem, the weapon system, characterized by the vector $z = (z_1, \cdots, z_n)$, where z_i ($\equiv y_i/K$ in notation of Section 1) is the proportion of money spent on weapons of the i-th kind.

We shall assume that V can take only two values, 1 (victory) and 0 (no victory), and that the utility function

$$
\begin{aligned}
w \equiv w(V, K) &\equiv w_0(K) = 0 \quad \text{for} \quad V = 0 \\
&\equiv w_1(K) \geq 0 \quad \text{for} \quad V = 1 .
\end{aligned}
\tag{3.1}
$$

[We leave to the reader to work out the meaning and the consequences of relaxing those assumptions.] Thus $w_1(K)$ is the utility of having victory while being poorer by K dollars. If the probability of victory were an exact function of K and z, say g(K, z), then the expectation $W \equiv Ew$ of social utility would be exactly $w_1(K) \cdot g(K, z)$. Actually, the probability of victory depends on the number x of targets destroyed. As in Section 2, this probability is identical with military utility u(x) and is, like x itself, a random variable. Let the distribution density of x, for given K and z, be f(x|K, z). The expectation W of social utility is obtained by integrating over x:

$$W \equiv \int_x w_1(K)u(x)f(x|K, z)dx = w_1(K) \cdot U(K, z), \tag{3.2}$$

where U denotes, as before, the expected military utility:

$$U(K, z) \equiv \int u(x)f(x|K, z)dx. \tag{3.3}$$

We see that the social maximand (3.2) depends on the strategic (i.e., controlled) variables K, z in a fashion that is determined by three functions: w_1, u, and f. Of these, the function u is described by parameters such as α, β in (2.1) that are approximately known. A function f will be derived in Section 4 from considerations on the performance of weapons. As to the function w_1, the "utility function of material wealth in the case of victory," it is not known with any precision except that, necessarily, $w_1(K)$ decreases as the expenditure K increases. If U(K, z) is monotone in K for given z, we can derive the "opportunity curve" ("efficient set") associating K with, for any fixed K, the highest achievable value of U—and therefore of W. Figs. 5, 7, 9, and 11 below give examples of such opportunity curves. Figs. 6, 8 and 10 answer the question: "If the expenditure is increased from K_0 to $K_0 + 1$ dollars, and the best weapon system is always used, by how many per cent will the expected probability of victory increase?" The policy maker who has this question of his answered, can then also answer for himself the question: "Under these circumstances, does it pay for the nation to increase its expenditure from K_0 to $K_0 + 1$ dollars?" That is, he will be forced to trace his welfare function w(V, K), or at least the relevant segment of it. If he is "rational," such a function must exist (for reasons analogous to those stated in Section 2), although it takes calm deliberation to state it.

We conclude this "ethical" Section by a remark on the "cost in lives." It has been argued convincingly that, in discussing offensive weapons, the cost of lives of the crew should not be regarded as a "final good": a costly mechanization of weapons aiming solely at economizing on lives of the crew may imply the waste of other lives (military or civilian) due to the withdrawal of resources from defense or from other forms of attack. But the cost of lives can and must be discussed if the over-all defense-and-attack strategy is discussed. The Jeffersonian triad of life, liberty (which would be lost through defeat) and the pursuit of happiness (achieved through consumption) as the ultimate "goods" is expressed by generalizing the social utility (welfare) function of (3.1) into

$$\begin{aligned} w \equiv w(V, L, K) &= w_0(L, K) = 0 \text{ for } V = 0 \\ &= w_1(L, K) \geq 0 \text{ for } V = 1, \end{aligned} \tag{3.4}$$

where L stands for "lives." Even if this special assumption is granted, there remains a major difficulty that arises from the fact that, unlike the budget K, the number of lives L is not a controlled variable. We cannot decide upon a given expenditure of lives. L is a random variable, and its distribution is, in general, a joint one with the number of targets: the joint density f(x, L), say, depends on the controlled variables K, z. The expected social utility takes the form

$$W = \int_x \int_L w_1(L, K) \cdot u(x) f(x, L|K, z) dx\, dL. \tag{3.5}$$

In this case, as distinct from the previous case, where "lives" were not discussed, we are not able to factor out the utility function w_1 and thus to construct the efficient set, if we know only that $w_1(L, K)$ is monotone decreasing in both L and K. Additional assumptions on the form of $w_1(L, K)$ are needed. For example, one might devise assumptions that would enable us to use the expected loss of lives as a controlled variable, and to construct the efficient set in the space whose coordinates are: the budget K, the expected probability of victory U, and the expected loss of lives. Such an assumption would probably imply that social utility w is linear in lives, and this is hardly plausible.[3] We shall not pursue this matter further in the present paper, and shall confine our considerations to the expected probability of victory (U) that can be bought for a fixed amount of money (K).

4. PERFORMANCE AND EXPECTED MILITARY UTILITY OF A WEAPON SYSTEM

In Section 1 we have defined y_i (i = 1, ···, n), the number of weapon units of the i-th kind. The vector $y \equiv (y_1, \cdots, y_n)$ defines a weapon system. We shall conceive a weapon system as a stockpile to be used in a campaign of a fixed duration. The performance of a weapon system y is defined as the probability density function f(x|y) which, for any x, gives the probability that x targets will be destroyed. By the performance of the i-th kind of weapon we mean the probability p_i that one weapon of this kind, when sent against a given target, will destroy it. In this paper, we shall make assumptions that will permit a particularly simple derivation of f(x|y) from a given set of p_i's.

We shall assume that the performance of each individual weapon unit is independent of the action of other weapons, whether of the same or of different kind. We are aware that this means neglecting the area defense and the "saturation." We believe, however, that by thinking through the implications of such simplifying assumptions we shall contribute to clarifying the matter and thus prepare the way for a more complete study. See also Section 8 below.

Roughly, one may define as a weapon unit of a given kind, one cell consisting of a certain number of guided missiles of a given type, carrying a certain number of bombs with specific characteristics; or one take-off of a cell of planes, similarly characterized.[4]

It will be remembered that we have assumed (in Section 1) all targets to be identical with respect to the strength of their local defense. Hence, for a single weapon unit, the

---

[3] The usual efforts of the defender to disperse his people testify to this. When the enemy attacker is too far or too high to aim at a particular crowd, dispersal affects only the variance, not the expectation of casualties. That attempts are made to diminish the variance shows that the social utility function is usually strictly concave, not linear, with respect to the number of casualties; see Section 2.

[4] We are indebted to A. Alchian and M. Davies for their explanations.

probability of getting to the target depends on the properties of that weapon only. If we multiply this probability by the probability of hitting the target (which, in turn, depends on the probable aiming error and the lethal radius of the weapon), and by a reliability coefficient (which depends mainly on the probability of misfiring), we obtain p_i, the probability that one weapon unit of the i-th kind will destroy its assigned target.

The performance of each weapon unit being assumed to be independent of the presence of other weapon units of the same kind, the probability that y_i units of the i-th kind destroy x_i targets is given by the term

$$\binom{y_i}{x_i} p_i^{x_i} q_i^{y_i - x_i}, \qquad q_i \equiv 1 - p_i, \tag{4.1}$$

of a binomial distribution. The mean and the variance of this distribution are, respectively,

$$\begin{aligned} \mu_i &\equiv p_i y_i \\ \sigma_i^2 &\equiv p_i q_i y_i \end{aligned} \tag{4.2}$$

The numerical magnitudes involved are such as to permit the approximation of the binomial distribution (4.1) by a normal distribution whose mean and variance are μ_i and σ_i as defined in (4.2).

We are interested in the distribution of the total number $x = \Sigma x_i$ of targets destroyed. Since each x_i is distributed (approximately) normally, and since it was assumed that the performance of each weapon unit of a given kind is independent of the presence of other weapon units, x has (approximately) normal distribution, whose mean and variance are, respectively,

$$\begin{aligned} \mu &= \Sigma \mu_i = \Sigma p_i y_i \\ \sigma^2 &= \Sigma \sigma_i^2 = \Sigma p_i q_i y_i \end{aligned}, \tag{4.3}$$

the summation being from 1 through n.

The expected military utility, defined in (1.1), becomes

$$U = \int_{-\infty}^{\infty} u(x) \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} dx; \tag{4.4}$$

by substituting $t = \frac{x - \mu}{\sigma}$,

$$U = \int_{-\infty}^{\infty} u(\sigma t + \mu) \cdot \phi(t) dt, \tag{4.5}$$

where $\phi(t)$ is the normal distribution density with zero mean and unit variance:

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

U is thus a function of μ and σ, which are in turn functions of the vector y. U is to be maximized by a proper choice of the weapon system y, given the constraints (1.2), (1.3).

We shall assume u(x) to be non-constant and non-decreasing (i.e., we neglect the possibility mentioned in Section 2). It follows from this assumption and (4.5) that U is strictly monotone increasing in μ for a given σ. This proposition will be used in Section 6 below.

If the special form (2.2) of the military utility function is assumed,

$$u(x) = 0, \quad x < \beta$$

$$= 1, \quad x \geq \beta,$$

we have, by (4.5),

$$U = \int_{\frac{\beta - \mu}{\sigma}}^{\infty} \phi(t)dt = \int_{-\infty}^{\frac{\mu - \beta}{\sigma}} \phi(t)dt .$$

Therefore, U is a monotone increasing function of $(\mu - \beta)/\sigma$, so that maximizing U is equivalent to maximizing

$$\omega = \frac{\mu - \beta}{\sigma}.$$

Note also that, of course, as in the general case, U is monotone increasing in μ for fixed σ. Moreover, if $\mu < \beta$, $U < \frac{1}{2}$ and is monotone increasing with σ; if $\mu > \beta$, $U > \frac{1}{2}$ and is monotone decreasing with σ. This confirms some considerations made in Section 2. It follows that for low levels of budget it may be preferable to concentrate on weapons with higher variance (i.e., with higher $p_i q_i$ and therefore with p_i near $\frac{1}{2}$), other things being equal. This will be confirmed in the numerical illustrations of Section 10, where the somewhat more general utility function (2.1) will also be used.

5. COST OF A WEAPON SYSTEM

In general, we have the cost constraint (1.3):

$$K = C(y_1, \cdots, y_n), \tag{5.1}$$

a general cost function which takes account of interdependence of various kinds of weapons

as joint products. An important class is formed by concave cost functions (defined in Section 7) which represent the case of "economies of large scale production and operation." A somewhat more restricted function than (5.1) is

$$K = \Sigma c_i(y_i); \tag{5.2}$$

here the interdependence of weapon costs is neglected: each weapon has its own cost function which is independent of the output of other weapons. A still further simplification is to neglect increasing or decreasing average costs for each weapon and to assume

$$K = \Sigma c_i y_i \ , \tag{5.3}$$

where each c_i is a constant.

6. CONSEQUENCES OF LINEAR COST AND INDEPENDENT PERFORMANCE

In Section 4, the assumption of independent performance of weapons was made, and was shown to imply that: (1) the number of targets destroyed is approximately normally distributed, its mean μ and its variance σ^2 being linear combinations (4.3) of the numbers y_i of the weapons of different kinds; (2) the expected military utility U is, by (4.5), monotone increasing in μ for any given σ. In the present Section we shall preserve the assumption of independent performance, combining it with various simple assumptions on the form of the utility function u(x) and of the cost function C(y).

In Section 2 we discussed a special case: a military utility function linear (and increasing) in the number of targets. It is equivalent to the requirement that the expected number of targets be maximized, given the amount of money available. If we could accept this linear utility function as well as the linear cost function (5.3), the solution would be particularly simple. We should have, then, to maximize the expected number of targets

$$\mu = \Sigma p_i y_i$$

with respect to the vector $(y_1, \cdots, y_n)$, subject to

$$\Sigma c_i y_i = K, \qquad y_i \geq 0, \qquad i = 1, \cdots, n.$$

We can, instead, maximize μ/K, i.e., the weighted average

$$\Sigma \frac{p_i}{c_i} z_i \ , \tag{6.1}$$

the weights z_i being defined as the proportions of K spent on each weapon:

$$z_i \equiv \frac{c_i y_i}{K} \geq 0 , \qquad \Sigma z_i = 1 . \tag{6.2}$$

Obviously (6.1) is largest when weight $z_j = 1$ is assigned to that weapon j for which the ratio

p_j/c_j ("probability per dollar") is larger than for any other weapon. That is, all money is spent on the j-th weapon.[5]

For n = 2, the case may be illustrated with the help of indifference lines (Fig. 3) in the (y_1, y_2)-plane (compare Figs. 12 and 13).

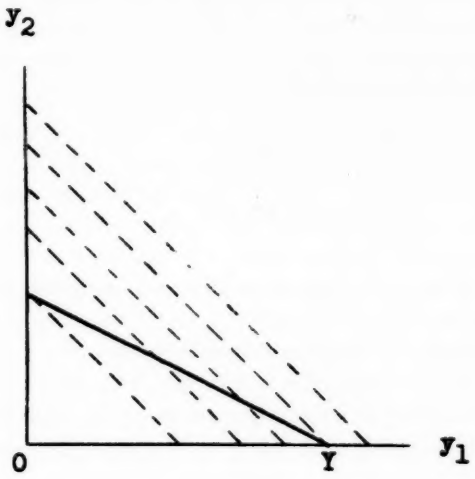

Figure 3

Each of the parallel straight lines represents the equation $\mu = \Sigma p_i y_i$ for a different value of μ. The thick single line represents the constraint $\Sigma c_i y_i = K$. Since the y_i are nonnegative, the highest value of μ is obtained at the "corner," Y, with $y_2 = 0$ and $y_1 = K/c_1$.

Thus, if independence of performances is assumed along with linear cost function and linear utility function, the best weapon system will consist of a single kind of weapon. More strictly, we must provide also for the case when there exists a set of two (or more) weapons for which the probability per dollar is the same, and is larger than for any weapon not in the set. Any weapon in that set and any "mixture" of those weapons is an optimal weapon system. Thus if utility function is linear, there will always exist a solution which is a one-weapon system (a "pure" solution).

If we maintain the utility function to be monotone nondecreasing but make it non-linear (as in most of Section 2), and if we still maintain the assumption of independent performance of weapons and of a linear cost function, we can show that there will always exist a solution that is a mixture of at most two weapons. This follows from Section 4, because of the "Theorem of Mixtures" that will be briefly stated and loosely proved later in this section.

It was shown in Section 4 that if weapon performances are independent, and if the utility function is non-decreasing, then U is strictly monotone increasing in μ for a given σ,

$$U = F(\mu, \sigma^2),$$

[5] An economist may think of the following analogy. Interpret y_i as the number of tons of the i-th fuel, c_i as its cost per ton, and p_i as the energy content per ton. Then K is the total cost, and μ the total energy content, and it is best to invest all money into the fuel with the highest energy content per dollar, p_i/c_i.

say. With cost function linear, we can write

$$\mu = \Sigma\, a_i z_i$$

$$\sigma^2 = \Sigma\, b_i z_i$$

where z_i is as defined in (6.2), and where

$$a_i = Kp_i/c_i\,,$$

$$b_i = Kp_iq_i/c_i\,.$$

That is, a_i and b_i are, respectively, the expectation and the variance of damage if all money is spent on the i-th weapon.

The "Theorem on Mixtures" can be stated as follows:

> Let $A = \Sigma a_i z_i$ and $B = \Sigma b_i z_i$, $\Sigma z_i = 1$, $z_i \geq 0$, $i = 1, \cdots, n$. Let $F(A, B)$ be monotone increasing in A for any given B. Then there exist non-negative z_i, z_j $(z_i + z_j = 1)$ such that F has its maximum when $A = a_i z_i + a_j z_j$, $B = b_i z_i + b_j z_j$.

(In our case, μ and σ^2 play, respectively, the roles of A and B, while all other notations remain unchanged.)

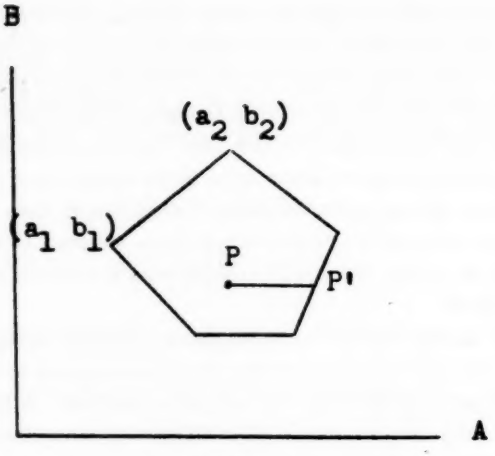

Figure 4

A neat geometric proof was outlined by J. Wolfowitz. On Fig. 4, each pair (a_1, b_1), (a_2, b_2), $\cdots$, (a_n, b_n) is represented by a point. The smallest convex polygon containing these points will also contain all possible linear combinations $A = \Sigma a_i z_i$, $B = \Sigma b_i z_i$ since the z_i are nonnegative numbers adding up to unity. In particular, the vertices of the polygon represent "pure" cases (with $z_i = 1$ for a certain i, and = 0 for all others), and its sides represent "mixtures of two." Draw a horizontal line through any point P belonging to the polygon. All points on this line that belong to the polygon have the same value of B; of all

these points, the one (P') with the largest value of A will lie on the right-hand boundary of the polygon. Therefore, the point that maximizes F will lie on the boundary, i.e., on a side or in a vertex, of the polygon.

Thus, if weapon performances are independent, the cost functions linear, and the utility function non-decreasing, it suffices to find, for any value of K, the one weapon or possibly the pair of weapons that constitute an optimal system. In Section 9, the problem is solved for n = 2, assuming the utility function stated in (2.2) (a step-function). This solution is applied, in Section 10, to find the optimal weapon system for varying values of β and K, when n = 2 and n = 4; to compute the expected utility of a given budget when the optimal weapon system is used; and to compute the marginal expected utility of a dollar at a given budget level, when optimal weapon systems are used. A more general utility function, that of (2.1), is used as a further illustration. The performance and cost parameters (ρ_i, c_i) were chosen for illustrative purposes only, but a systematic numerical study of the results, for an appropriate grid of values of ρ_i, c_i, can be made if necessary.

As will be shown in the next section, the "mixture theorem" applies also to the case when the average costs are not constant, but, more generally, the cost function is concave, because of advantages of large scale production and operation. However, the actual solutions worked out so far apply to the linear case only (Section 9 and Section 10).

7. EXTENSION TO CONCAVE COST FUNCTIONS (LARGE SCALE ECONOMIES)

The constant average cost condition was reflected in the mixture theorem of Section 6 by the condition $\Sigma z_i = 1$, which was a normalization of the condition $\Sigma c_i y_i = K$. In the present section we will replace this by the cost constraint $C(y) \leq K$ where $C(y)$ is concave. For convenience we will introduce the restriction that the coefficients a_i and b_i of $A = \Sigma a_i y_i$, $B = \Sigma b_i y_i$ be positive. With these changes the theorem of Section 6 becomes:

Let F(A, B) be monotone increasing in A for every B in the domain defined by

$$A(y) = \Sigma a_i y_i \qquad a_i > 0$$

$$B(y) = \Sigma b_i y_i \qquad b_i > 0 \qquad i = 1, \cdots, n \tag{7.1}$$

$$y_i > 0, \qquad C(y) \leq K. \tag{7.2}$$

Let C(y) satisfy the inequality

$$C[(1-\alpha)y_0 + \alpha y_1] \geq (1-\alpha)C(y_0) + \alpha C(y_1) \tag{7.3}$$

for every y_0, y_1 with nonnegative components and every α in the interval $0 \leq \alpha \leq 1$.

Then, if F(A, B) has a maximum in the above domain, it is attained for some y such that at most two components are positive.

PROOF: It is sufficient to show that for every y_0 satisfying the constraints (7.2) there exists a y_1 having not more than two positive components such that y_1 satisfies (7.2) and

$$A(y_1) = A(y_0)$$

$$B(y_1) = B(y_0).$$

Let y_0 satisfying (7.2) be given. Then, by virtue of the positiveness of the a_i, b_i, the set of y, $y_i \geq 0$, such that

$$A(y) = A(y_0)$$

is a convex polygon. Similarly, the set of y such that

$$B(y) = B(y_0)$$

is a convex polygon. Since the intersection is not empty, it too is a convex polygon, and is spanned by a set of vectors $v_1, \cdots, v_m$ such that each vector of the set has not more than two positive components. Since y_0 lies in the intersection, there exist numbers $\alpha_i \geq 0$, $\sum_{i=1}^{m} \alpha_i = 1$, such that

$$y_0 = \sum_{i=1}^{m} \alpha_i v_i .$$

From (7.3) it follows that

$$\Sigma \alpha_i C(v_i) \leq C(y_0) \leq K ,$$

so that for at least one v_i, v_j, say,

$$C(v_j) \leq K .$$

Therefore, v_j satisfies the constraints (7.2) and the equalities

$$A(v_j) = A(y_0)$$

$$B(v_j) = B(y_0).$$

By letting y_0 be a vector for which the maximum F(A, B) is attained, the theorem follows.

8. EXTENSION TO A CASE OF NON-INDEPENDENT PERFORMANCES

That the rather severe assumption of independence of weapon unit performance is not essential to the general methods we have employed is shown by the following construction of a weapon system performance. Let p_i be the conditional probability that a weapon unit of the i-th kind destroys a target, given that the weapon unit reached the target, and let $q_i(y)$ be the probability that a weapon unit of the i-th kind reaches the target. The $q_i(y)$ are thus functions of the weapon system used. The probability distribution of the number of targets destroyed by weapons of the i-th kind, x_i, is then (dropping subscripts on x_i and y_i for convenience)

$$f_i(x|y) = \sum_{z=x}^{y} \binom{y}{z} p_i^z (1 - p_i)^{y-z} \binom{z}{x} [q_i(y)]^x [1 - q_i(y)]^{z-x}$$

$$= \binom{y}{x} [p_i q_i(y)]^x \sum_{j=0}^{y-x} \binom{y-x}{j} (p_i[1 - q_i(y)])^j (1 - p_i)^{y-x-j}$$

$$= \binom{y}{x} [p_i q_i(y)]^x [1 - p_i q_i(y)]^{y-x}.$$

Since this distribution is binomial, much of the discussion of Section 4 applies to this model equally as well as it does to the model of that section. In making the normal approximation we have

$$\mu = \Sigma\, y_i p_i q_i(y)$$

$$\sigma^2 = \Sigma\, y_i p_i q_i(y)\,[1 - p_i q_i(y)]\,.$$

Since μ and σ^2 are no longer linear in the weapon vectors, the mixture theorems do not apply, which increases the computational difficulties of explicitly finding optimal weapon systems. See Section 11 for a fuller discussion of the point.

9. SOLUTION FOR THE CASE OF UTILITY FUNCTION (2.2) WHEN n = 2

As was noted in Section 4, if the special form (2.2) of the military utility function is assumed, maximizing

$$u = \int_{-\infty}^{\frac{\mu - \beta}{\sigma}} \phi(t)dt$$

is equivalent to maximizing $\omega = (\mu - \beta)/\sigma$. By Section 6 it suffices to solve the problem for the case n = 2. Following the notation of Section 6,

$$\mu = a_0(1 - z) + a_1 z = a_0 + (a_1 - a_0)z$$

$$\sigma^2 = b_0(1 - z) + b_1 z = b_0 + (b_1 - b_0)z \qquad 0 \le z \le 1$$

so that the problem of maximizing

$$\omega = \frac{a_0 + (a_1 - a_0)z - \beta}{\sqrt{b_0 + (b_1 - b_0)z}} \qquad 0 \le z \le 1$$

is readily solved by elementary methods.

Although the problem is solved by elementary methods, it is of interest to consider conditions on the parameters which yield a mixture of weapons, i.e., an interior solution for z. Appropriate inequalities are readily derived from the necessary and sufficient conditions:

$$\left.\frac{d\omega}{dz}\right|_{z=0} = \frac{1}{\sqrt{b_0}}\left[a_1 - a_0 - \frac{(a_0 - \beta)(b_1 - b_0)}{2b_0}\right] > 0$$

(9.1)

$$\left.\frac{d\omega}{dz}\right|_{z=1} = \frac{1}{\sqrt{b_1}}\left[a_1 - a_0 - \frac{(a_1 - \beta)(b_1 - b_0)}{2b_1}\right] < 0.$$

These inequalities may be combined to give

$$\text{(9.2)} \qquad \frac{(a_0 - \beta)(b_1 - b_0)}{2b_0} < a_1 - a_0 < \frac{(a_1 - \beta)(b_1 - b_0)}{2b_1}$$

Since we may, without loss of generality, take $b_1 > b_0$, the above may be put in the more convenient form:

$$\text{(9.3)} \qquad a_0 - 2b_0\frac{a_1 - a_0}{b_1 - b_0} < \beta < a_1 - 2b_1\frac{a_1 - a_0}{b_1 - b_0} = a_0 - 2b_0\frac{a_1 - a_0}{b_1 - b_0} - (a_1 - a_0).$$

From (9.3) we immediately obtain the necessary condition:

$$(b_1 - b_0)(a_1 - a_0) < 0.$$

In terms of the initially given parameters this yields

$$\text{(9.4)} \qquad \left(\frac{q_1}{q_0}\frac{p_1}{c_1} - \frac{p_0}{c_0}\right)\left(\frac{p_1}{c_1} - \frac{p_0}{c_0}\right) < 0.$$

We also obtain the necessary condition:

$$\max(a_0, a_1) < \beta$$

or

$$\text{(9.5)} \qquad \max\left(\frac{p_0}{c_0}, \frac{p_1}{c_1}\right) < \frac{\beta}{K}.$$

Note that in the case of n weapons, if K and β are such that $p_i/c_i > \beta/K$ holds for at least one i, we choose the weapon for which

$$\frac{\frac{p_i}{c_i} - \frac{\beta}{K}}{\sqrt{\frac{p_i q_i}{c_i}}}$$

is maximized. This is not necessarily the weapon which yields the largest expected number of targets killed.

For completeness we give the solution to the maximizing problem given at the beginning of this section. In order to emphasize the dependence on K and β, the following notation is convenient. Let

$$r_i = \frac{p_i}{c_i}, \quad s_i = \frac{p_i q_i}{c_i}, \quad \Delta r = r_1 - r_0, \quad \Delta s = s_1 - s_0,$$

so that $a_i = Kr_i$, $b_i = Ks_i$. The following assumes $\Delta s > 0$. If $\Delta r > 0$, then no "mixture" solution exists and we have:

(9.6)

$$\max \omega = \sqrt{K}\frac{r_0 - \frac{\beta}{K}}{\sqrt{s_0}}, \qquad \frac{\beta}{K} \le r_0 + \frac{\Delta r}{1 - \sqrt{\frac{s_1}{s_0}}};$$

$$= \sqrt{K}\frac{r_1 - \frac{\beta}{K}}{\sqrt{s_1}}, \qquad \frac{\beta}{K} \ge r_0 + \frac{\Delta r}{1 - \sqrt{\frac{s_1}{s_0}}}.$$

If $\Delta r < 0$, mixture solutions are possible for appropriate values of K and β.

$$\max \omega = \sqrt{K}\frac{r_0 - \frac{\beta}{K}}{\sqrt{s_0}}, \qquad \frac{\beta}{K} \le r_0 - \frac{2s_0\Delta r}{\Delta s};$$

(9.7)

$$= 2\sqrt{K}\frac{r_0 - \frac{\beta}{K} - \frac{s_0\Delta r}{\Delta s}}{\sqrt{(r_0 - \frac{\beta}{K})\frac{\Delta s}{\Delta r} - s_0}}, \qquad r_0 - \frac{2s_0\Delta r}{\Delta s} \le \frac{\beta}{K} \le r_0 - \frac{2s_0\Delta r}{\Delta s} - \Delta r;$$

$$= \sqrt{K}\frac{r_1 - \frac{\beta}{K}}{\sqrt{s_1}}, \qquad \frac{\beta}{K} \ge r_0 - \frac{2s_0\Delta r}{\Delta s} - \Delta r.$$

The interior maximum is attained at

$$z = \frac{r_0 - \frac{\beta}{K}}{\Delta r} - \frac{2s_0}{\Delta s}. \tag{9.8}$$

10. NUMERICAL ILLUSTRATIONS FOR THE CASE OF UTILITY FUNCTION (2.2) WITH n = 2 AND n = 4; AND FOR UTILITY FUNCTION (2.1) WITH n = 2

Three examples are given as illustrations of the solution of the weapon selection problem for utility functions of type (2.2). Example 1 illustrates a case of two weapons in which mixture solutions occur. Example 2 is concerned with the two-weapon problem for the case in which the best use of any budget is to expend it entirely on a single weapon.

Example 3 illustrates the choice problem for four weapons for the case in which mixed solutions occur for appropriate budget values.

In Examples 1 and 2 three utility functions are considered, specified by the values $\beta = 50, 75, 100$. In Example 3 a single utility function, $\beta = 50$, is considered.

The solutions will be indicated by specifying the ranges in which the entire budget is expended on a single weapon, and, in the case of mixtures, by specifying the budget value for which half of the budget is expended on each weapon of the mixture. The maximum expected utilities and the marginal maximum utilities are plotted in Figures 5 to 10.

Examples 1 to 3 are characterized by the following tables.

EXAMPLE 1

Table 1. Weapon Characteristics, Example 1.

Weapon	Kill Probability	Cost
A	0.4	1.00
B	0.8	1.85

The maximum expected military utility (the utility of the best feasible weapon system) is plotted in Fig. 5 as a function of the budget available. Fig. 6 shows the derivative of this function, the so-called "marginal utility" of the budget (or utility of an additional dollar), assuming that the budget is expended on the best feasible weapon system.

Table 2. Weapon Choice, Example 1.

Weapon Choice	Budget K		
	$\beta = 50$	$\beta = 75$	$\beta = 100$
A	$K \leq 99.74$	$K \leq 149.61$	$K \leq 199.48$
mix A and B 50-50	$K = 103.07$	$K = 154.61$	$K = 206.14$
B	$K \geq 106.63$	$K \geq 159.95$	$K \geq 213.27$

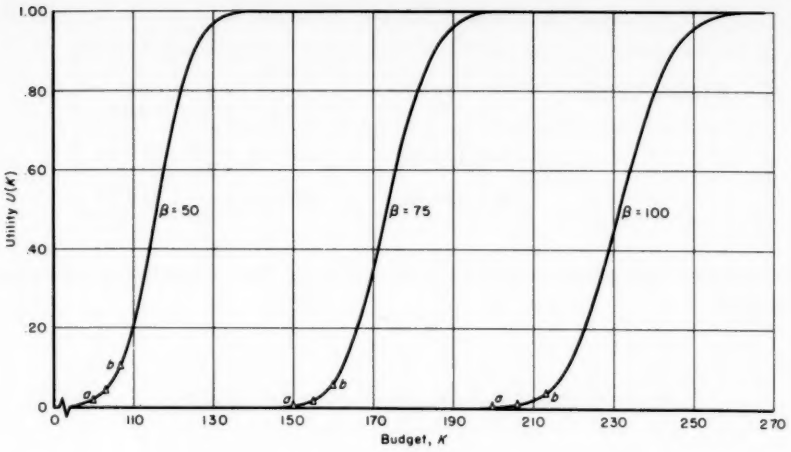

Figure 5 - Maximum expected military utility as a function of budget, example 1.

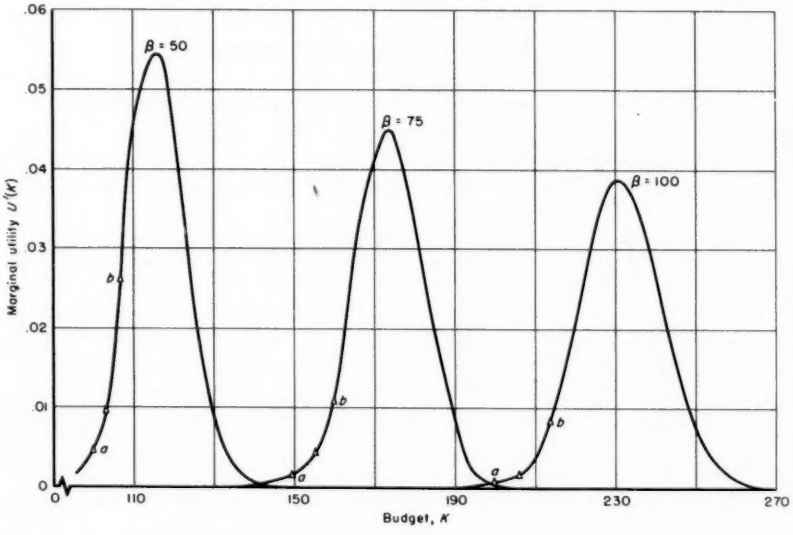

Figure 6 - Marginal utility as a function of budget, example 1.

EXAMPLE 2

Table 3. Weapon Characteristics, Example 2.

Weapon	Kill Probability	Cost
A	0.4	1
B	0.8	2.10

Table 4. Weapon Choice, Example 2.

Weapon Choice	Budget K		
	$\beta = 50$	$\beta = 75$	$\beta = 100$
A	$K \leq 139.66$	$K \leq 209.49$	$K \leq 279.31$
B	$K \geq 139.66$	$K \geq 209.49$	$K \geq 279.31$

The maximum expected military utility is plotted in Fig. 7 and the marginal utility is plotted in Fig. 8.

EXAMPLE 3

Table 5. Weapon Characteristics, Example 3.

Weapon	Kill Probability	Cost
A	0.2	1.0
B	0.4	1.9
C	0.6	2.8
D	0.8	3.7

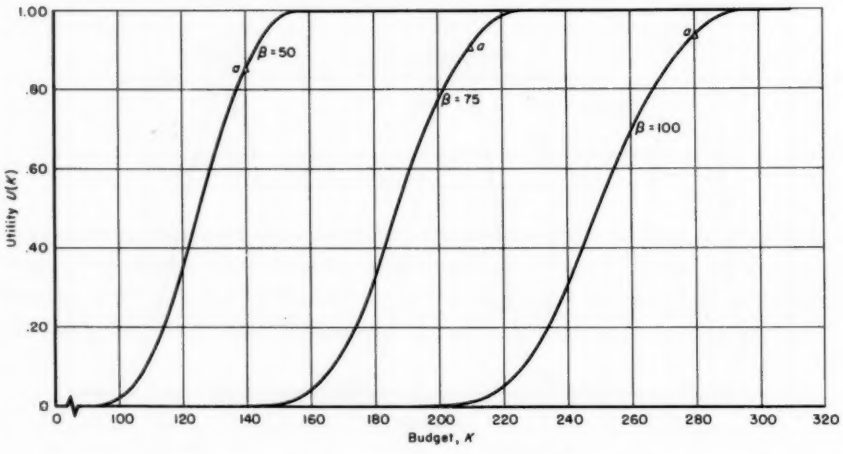

Figure 7 - Maximum expected military utility as a function of budget, example 2.

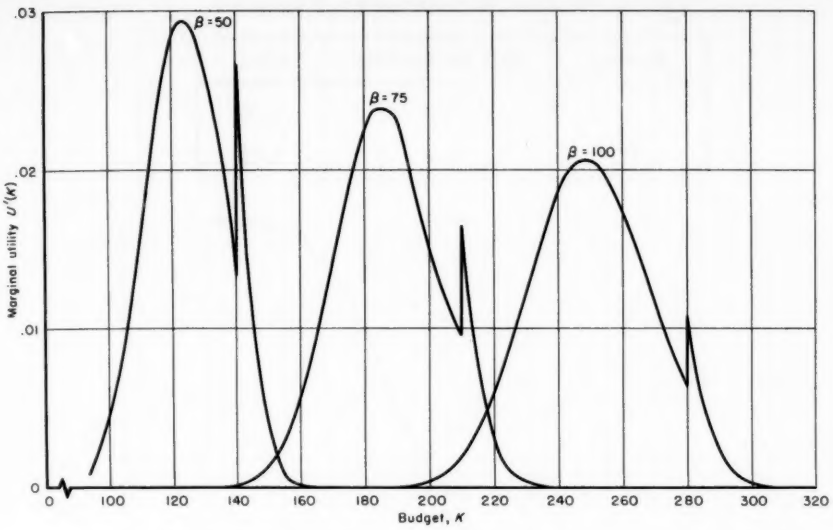

Figure 8 - Marginal utility as a function of budget, example 2.

Table 6. Weapon Choice, Example 3.

Weapon Choice	Budget
A	$K \leq 166.83$
mix A and B 50-50	$K = 169.81$
B	$172.89 \leq K \leq 213.58$
mix B and C 50-50	$K = 215.33$
C	$217.11 \leq K \leq 225.23$
mix C and D 50-50	$K = 226.19$
D	$K \geq 227.17$

The maximum expected utility is plotted in Fig. 9 and the marginal utility is plotted in Fig. 10.

EXAMPLE 4

Example 4 is concerned with a utility function of type (2.1). The utility function is specified by the values $\alpha = 25$, $\beta = 75$.

$$u(x) = 0, \qquad x \leq 25$$

$$= \frac{x - 25}{50}, \qquad 25 \leq x \leq 75$$

$$= 1, \qquad 75 \leq x \quad .$$

The weapons considered are taken from Example 1.

Table 7. Weapon Characteristics, Example 4.

Weapon	Kill Probability	Cost
A	0.4	1.00
B	0.8	1.85

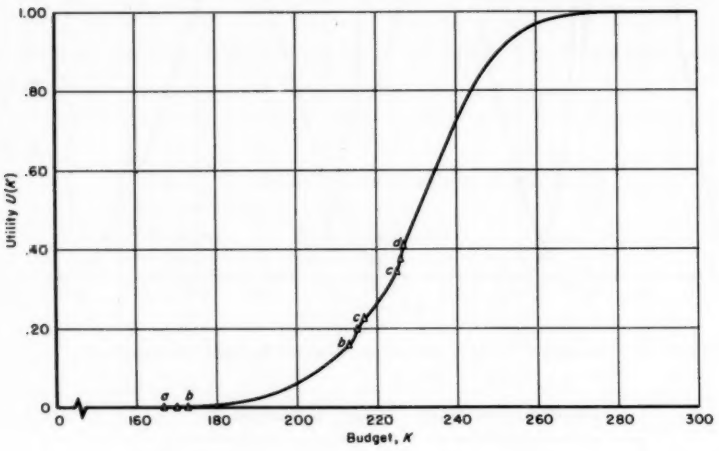

Figure 9 - Maximum expected military utility as a function of budget, example 3. ($\beta = 50$)

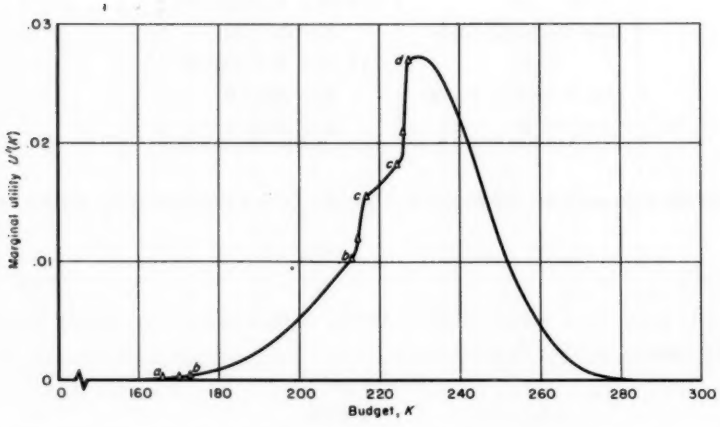

Figure 10 - Marginal utility as a function of budget, example 3.

Since the choice problem was solved numerically and it was not felt necessary to attain great accuracy, the interval of budget values yielding mixture solutions was determined only approximately, roughly to within one unit of budget. The calculation of the marginal maximum utility is quite tedious in the budget interval of mixtures, and only the curve of maximum utility is plotted. This curve is given in Fig. 11. The solution is indicated in Table 8.

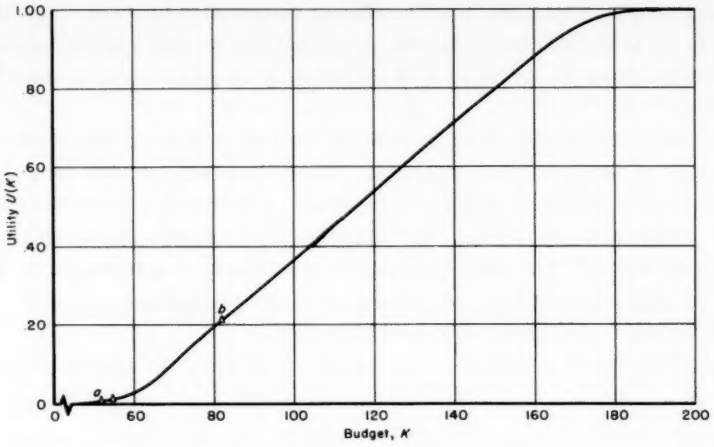

Figure 11 - Maximum expected military utility as a function of budget, example 4.

Table 8. Weapon Choice, Example 4.

Weapon Choice	Budget
A	$K \leq 52.5$
mix A and B 50-50	$K = 55$
B	$K \geq 81$

11. DESIRABLE GENERALIZATIONS; NON-LINEAR PROGRAMMING

It was possible, in Sections 9 and 10, to apply elementary methods to solve the problem of the optimal weapon system. This was due to three assumptions:

(1) non-decreasing utility function u(x);

(2) constant average cost, c_i, for each kind of weapon;

(3) independence of the performance (p_i) of each unit of weapon.

The first of these assumptions is not too restrictive, but the second and third should be removed if any more complete study of the problem is attempted.

In Section 6, utility function u(x) was linear, and this, together with the assumptions (2) and (3), made our problem one in "linear programming": "maximize a linear function U(y) subject to a linear equality $\Sigma c_i y_i = K$ and to the non-negativity of the arguments y_i." In the rest of the paper, with u(x) non-linear, U(y) became also non-linear, and the problem became one in "non-linear programming." However, assumptions (2) and (3), which resulted in the conditions given in Section 6, reduced the problem to one of finding an optimal mixture of at most two kinds of weapons, a problem solvable by elementary methods. As was seen in Section 7, replacing (2) by the assumption that C(Y) is concave retains the elementary nature of the solution in that it is necessary to consider only the two-weapon problem. Further relaxing of (2) will most likely lose this simplifying property of the solution. If C(y) is convex, for example, optimal weapon systems consisting of only two types of weapons will be an exception rather than the rule. With the relaxation of (3) use of a normal approximation to the distribution of the number of targets destroyed may

become inappropriate, which will lead to difficult distribution problems. On the other hand, relaxing (3) in the manner of Section 8, so that the normal approximation is reasonable, raises difficulties due to the loss of linearity of the parameters μ and σ^2 in the weapon vectors.

It would seem worth-while to apply generalizations of assumptions (2) and (3), using general methods of non-linear programming. These have been studied by Kuhn and Tucker and by Slater. The problem of finding, in the small, maxima of a non-linear function of a vector, U(y), constrained by a non-linear equation, C(y) = 0, is a classical one, and is familiar to economists.[6] The essential additional difficulty is introduced by the presence of constraining inequalities (these are, in our case, the inequalities $y_i \geq 0$, $i = 1, \cdots, n$), and the requirement to find the maximum in the large.

In the classical case, the solution is obtained by forming the Lagrangian expression

$$L = U(y) + \lambda C(y) \tag{11.1}$$

and by requiring that

$$\partial L / \partial y_i = 0, \qquad i = 1, \cdots, n, \tag{11.2}$$

and that the leading principal minors (beginning with the third order) of the bordered matrix

$$\begin{pmatrix} 0 & \frac{\partial C}{\partial y_1} & \cdots & \frac{\partial C}{\partial y_n} \\ \frac{\partial C}{\partial y_1} & \frac{\partial^2 L}{\partial y_1^2} & \cdots & \frac{\partial^2 L}{\partial y_1 \partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial C}{\partial y_n} & \frac{\partial^2 L}{\partial y_1 \partial y_n} & \cdots & \frac{\partial^2 L}{\partial y_n^2} \end{pmatrix} \tag{11.3}$$

be alternatively positive and negative. This presupposes differentiability of U and C. The following geometrical interpretation is well known: In the y-space, condition (11.2) means that at the local optimum value $y = y_0$, the surface C(y) = 0 be tangent to a level surface ("indifference surface") $U(y) = U(y_0)$; while condition (11.3) means that, at $y = y_0$, the former of these two surfaces be less convex than the latter. Figures 12 and 13, analogous to Fig. 3, illustrate this. Dotted lines are "indifference lines," the thick line gives the constraint. Compare also Figure 3.

[6] See Hancock, Hicks, or Samuelson.

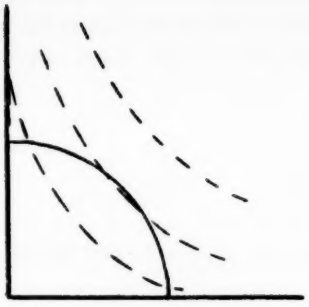

Figure 12

Figure 13

With the non-negativity condition introduced, $y_i \geq 0$, $i = 1, \cdots, n$, (11.2) and (11.3) become necessary conditions for a local interior maximum. For a maximum on the boundary, i.e., $y_i = 0$ for some i, Kuhn and Tucker gave some necessary conditions, involving the sign of the first derivative $\partial L/\partial y_i$ and the convexity of the functions C and U (not the level surfaces, or projections, of U), taken separately. Note that, on the contrary, condition (11.3) characterizes U and C taken jointly, viz., the "comparative convexity" of their level surfaces. In this sense, it is weaker than the Kuhn-Tucker condition, thus raising the question whether the latter could be weakened.[7]

To conclude, it seems that further study should be given to: (1) the way in which the performance of a weapon unit depends on the presence of other weapon units, and the resulting reformulation of the distribution $f(x|y)$ and the expected military utility function $U(y)$; (2) the general problem of non-linear programming, which may supply mathematical tools for the general problem of the optimal weapon system, and, in fact, for many other problems involving practical decisions.

The problem as studied in this paper is a special case of a more general one. It may not be possible to represent the weapon system y as a vector of quantities of several weapons; instead, we can regard y as an element of the set Y of all possible weapon systems (or, better, "military policies"—each policy being defined by certain offensive and defensive weapons, bases, etc.). The concept of "damage," x, can also be generalized: x may be interpreted as an element of the set X of all possible physical outcomes (each outcome being defined by various damages suffered by ourselves and by the enemy, by strategic gains of territory, etc.). Finally we define the set V of "military results," consisting (in the simplest case) of two elements only, v(= victory) and v'(= non-victory). We further define "performance" of the system y as the probability of x given y: $Pr(x|y)$. Let X consist of a finite number of elements (an assumption made here only for the sake of easier presentation) $x^{(1)}, x^{(2)}, \cdots$. We define as the utility of the j-th outcome, $u(x^j)$, the probability $Pr(v \mid x^j)$ that victory is obtained when the physical outcome is x^j. It follows that to every element y of Y there corresponds a quantity

$$\sum_j u(x^j) Pr(x^j|y) = U(y),$$

[7] This was pointed out by Slater, who gave weaker conditions and also dispensed with the requirement of differentiability. See also Arrow and Hurwicz.

which we call the expected utility of the weapon system y. In the present paper we have not pursued this general line of approach; however, most limitations of the results obtained are due to the fact that we had to take a narrower point of view, for the sake of simplicity and definiteness.

REFERENCES

[1] Arrow, K., and Hurwicz, W., "A Gradient Method for Approximating Saddle-points and Constrained Maxima," Rand Report P-223.

[2] Debreu, G., "Definite and Semi-definite Quadratic Forms," Econometrica, Vol. 20 (1952), pp. 295-300.

[3] Hancock, H., Theory of Maxima and Minima.

[4] Herstein, I. N., and Milnor, John, "An Axiomatic Approach to Measurable Utility," Econometrica, Vol. 21 (1953), pp. 291-297.

[5] Hicks, J. R., Value and Capital: Mathematical Appendix.

[6] Kuhn, H. W., and Tucker, A. W., "Non-linear Programming," Proceedings of the Second Symposium on Mathematical Statistics and Probability, Berkeley, 1950.

[7] Mann, H., "Quadratic Forms with Linear Constraints," American Mathematical Monthly, 1943.

[8] Marschak, J., "Rational Behavior, Uncertain Prospects, Measurable Utility," Econometrica, Vol. 18 (1950), pp. 111-141.

[9] v. Neumann, J., and Morgenstern, O., Theory of Games and Economic Behavior. Sec. 3 and Appendix to Second Edition.

[10] Samuelson, P., Foundation of Economic Analysis: Mathematical Appendix.

[11] Slater, M., "Lagrange Multipliers Revisited. A Contribution to Non-linear Programming," Cowles Commission Discussion Paper, Mathematics, No. 403.

ON SOME APPLICATIONS OF THE THEORY OF DYNAMIC PROGRAMMING TO LOGISTICS

Richard Bellman
The Rand Corporation

This article considers multi-stage processes composed of a sequence of operations in which the outcome of the preceding operations may be used to guide the course of future operations. Examples of such processes can be found in many fields including industrial production, military planning, inter-industry analyses, and learning processes. A mathematical theory of dynamic programming is developed to treat such multi-stage processes arising in the field of logistics with examples of problems amenable to the techniques developed.

1. INTRODUCTION

In recent years, the study of multi-stage processes has become of greater and greater importance in wider and wider fields. By a multi-stage process we mean a process composed of a sequence of operations in which the outcome of the preceding operations may be used to guide the course of future operations. Two types of operations may be distinguished immediately, those in which the outcome is completely determined, and those in which the outcome is uncertain. Inter-industry analysis, industrial production and military planning all furnish numerous examples of these processes, which occur also in such far-flung fields as the theory of learning processes and the study of mental breakdown.

To treat these problems in a systematic fashion we have developed the theory of dynamic programming. In this article, we shall consider some simple, but non-trivial, problems of the kind that plague logisticians, to illustrate what kinds of problems are amenable to our techniques, what analytic results may be expected, and what computational procedures must be utilized in general.

In order not to obscure the basic ideas with technical foliage, we have consistently attempted to pose the simplest version of a problem possessing any aspect of reality. The first problem we consider is a purchasing problem involving the ordering of different types of equipment with different salvage value. After formulating the problem in both new and old terms, we present a number of typical analytic results. Following this, we show how these results may be used as the basis of approximate solutions for more complicated problems.

Following this, we turn to a discussion of problems involving uncertainty in the result of our operation. Here the selection of an appropriate criterion, i.e. maximum expected outcome, probability of achieving at least a certain level, becomes a matter of great importance in determining optimal policies.

A typical problem involving uncertainty is the optimal inventory problem, first discussed by Arrow, Harris and Marschak, [2], and then in greater generality by Dvoretzky, Kiefer and Wolfowitz, [13].

Finally, to illustrate the application of a method of great utility and versatility, the method of continuous approximations, we shall consider a simple "smoothing problem."

A list of papers for those interested in further results and applications to related fields is given in the bibliography.

2. A PURCHASING PROBLEM

A general problem that recurs in logistics is that of purchasing equipment to perform an assigned task. It is usually true that the task is not a simple one, but rather composed of a number of component parts of diverse nature. Consequently, equipment which is ideally suited to one task may be mediocre in the performance of a second, and utterly worthless for a third. To overcome this, it may be necessary to purchase various kinds of equipment which will be blended to yield satisfactory performance.

Let us now consider three problems, of ascending order of realism, which contain various aspects of the above.

VERSION 1.

"Over a period of N years, it is necessary, at the beginning of each year, to order some equipment to perform certain tasks. We possess an initial quantity of money x, which is to be divided into two parts, y and x - y. The first amount, y, is to be used for the purchase of equipment of type A, and the remaining amount, x - y, is to be used for the purchase of equipment of type B. The y dollars spent for A-equipment will yield g(y) man-hours available for the assigned tasks during the first year, while the (x - y) dollars spent for B-equipment will provide h(x - y) man-hours. At the end of the year, the A-equipment may be salvaged to provide an equivalent of ay dollars for the purchase of new equipment, where $0 \leq a < 1$, and the B-equipment may be salvaged to provide an equivalent of b(x - y) dollars, where $0 \leq b < 1$. With this new total, ay + b(x - y), available for the purchase of new equipment, the process is repeated each year until the N-year period is over.

The problem is to determine the yearly allocation of money which will maximize the total man-hours obtained from the equipment over the N-year period."

This problem is intimately connected with the "phasing problem" arising in the replacement of old equipment by new. For a detailed discussion see Alchian, [1].

VERSION 2.

"In the same setting as above, we allow money to be held in reserve, which is to say it is not required to spend all the money available each year."

VERSION 3.

"In addition to the liberalization of policy introduced in Version 2, we assume that additional funds will be appropriated each year for the purchase of new equipment, r_k at the end of the k^{th} year."

In the next section we shall consider the conventional mathematical formulation of these problems, and point out the disadvantages of this approach.

3. THE CLASSICAL FORMULATION OF VERSION 1

Let $y_1, y_2, \ldots, y_N$ be the amounts of money used to purchase A-equipment at the beginning of the first, second, .., N^{th} years, respectively. The total man-hours obtained from both types of equipment over the N-year period will be

$$J(y_1, y_2, \ldots, y_N) = \sum_{i=1}^{N} g(y_i) + \sum_{i=1}^{N} h(x_i - y_i), \tag{3.1}$$

where

(3.2)
$$\text{(a)}\quad 0 \le y_i \le x_i$$
$$\text{(b)}\quad x_1 = x,$$
$$x_2 = ay_1 + b(x_1 - y_1)$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$x_N = ay_{N-1} + b(x_{N-1} - y_{N-1})$$

The problem enunciated above in Version 1 is equivalent to determining the maximum of $J(y_1, y_2, \ldots, y_N)$ over the N-dimensional region described by the inequalities in (3.2). Although calculus is useful, it must be applied with care, since several, and as we shall see below, in some cases, all, of the maxima may lie on the boundary of the region.

There is little difficulty in obtaining the answer computationally if N is small. However, for N of any size, say 10, the problem of searching N-dimensional space becomes quite onerous.

Observe, furthermore, that, as far as the practical man is concerned, this approach yields far too much information. Given x and N, there is no need to know $y_2, y_3, \ldots, y_N$. At the beginning of the first year, all that is required is y_1. Consequently, if we know y_1, as a function of x and N, the solution is complete.

Guided by this simple observation, we shall pursue an entirely different and novel approach in the next section.

4. THE DYNAMIC PROGRAMMING FORMULATION

It is readily seen that having fixed the functions g and h, the total number of man-hours obtained over an N-year period using an optimal policy of expenditures is a function only of x, the initial quantity of money, and N, the number of stages (years) remaining. Let us then define for N = 1, 2, ...,

(4.1) $f_N(x)$ = number of man-hours obtained over a period of N years, given an initial quantity of money x, and employing an optimal policy.

We have

(4.2)
$$f_1(x) = \max_{0 \le y \le x} [g(y) + h(x-y)].$$

Let us now consider the fundamental recurrence relations connecting $f_{N+1}(x)$ and $f_N(x)$. Considering the (N + 1)—stage process, let x be divided into y and x - y at the first year. The number of man-hours obtained over the (N + 1)—year span will be g(y) + h(x-y), those coming from the first year, plus the man-hours obtained over the remaining N-year period starting with an amount ay + b(x-y).

It is clear that regardless of the initial y-choice, the quantity ay + b(x-y) will be used in an optimal fashion over the remaining N-steps. Hence, the number of man-hours obtained from it

will be, by definition, $f_N(ay + b(x-y))$. Thus the total return due to an initial subdivision into y and (x-y) will be

$$R_{N+1}(x, y) = g(y) + h(x-y) + f_N(ay + b(x-y)). \tag{4.3}$$

Since we wish to maximize the total return, y must be chosen to maximize $R_{N+1}(x, y)$. This yields the equation

$$\begin{aligned} f_{N+1}(x) &= \max_{0 \le y \le x} R_{N+1}(x, y) \\ &= \max_{0 \le y \le x} [g(y) + h(x-y) + f_N(ay + b(x-y))], \end{aligned} \tag{4.4}$$

for N = 2, 3, ...

In subsequent sections we shall present some consequences of the result.

5. VERSION 2

Let us now derive the analogous system of equations for the second version of the problem. Let us asume, as is usual, that the value of money depreciates from year to year, so that x dollars now is worth cx next year, where $0 < c < 1$. Then, with the same definition of $f_N(x)$ as in Section 4, we have

$$f_{N+1}(x) = \max_{\substack{y_1 + y_2 \le x \\ y_1,\ y_2 \ge 0}} [g(y_1) + h(y_2) + f_N(ay_1 + by_2 + c(x-y_1-y_2))], \tag{5.1}$$

where y_1 is the amount allocated to the purchase of A-equipment and y_2 the amount allocated to the purchase of B-equipment.

6. VERSION 3

The third version is more complicated since the yearly appropriations will be, in general, different. Let

(6.1) r_k = additional appropriation for the purchase of new equipment received at the end of the k^{th} year,

$f_{N,k}(x)$ = number of man-hours obtained over an N-year period, starting from the beginning of the $(k+1)^{st}$ year, given an initial amount of money x and employing an optimal policy.

Arguing as above, we have

$$f_{N,k}(x) = \max_{\substack{y_1 + y_2 \le x \\ y_1,\ y_2 \ge 0}} [g(y_1) + h(y_2) + f_{N-1,k+1}(ay_1 + by_2 + c(x-y_1-y_2) + r_k)]. \tag{6.2}$$

7. INFINITE STAGE APPROXIMATION

A device of great usefulness in this and many similar problems in which the far-off future plays an insignificant role is that of assuming N to be infinite. This mathematical fiction introduces an essential simplification in the problem since the sequence $f_N(x)$ is replaced by a single function

$$f(x) = \lim_{N\to\infty} f_N(x). \tag{7.1}$$

In place of (4.4) we obtain the equation

$$f(x) = \max_{0 \le y \le x} [g(y) + h(x-y) + f(ay+b(x-y))]. \tag{7.2}$$

Similarly (5.1) becomes

$$f(x) = \max_{\substack{y_1,\ y_2 \ge 0 \\ y_1 + y_2 = x}} [g(y_1) + h(y_2) + f(ay_1+by_2+c(x-y_1-y_2))]. \tag{7.3}$$

If we attempt to use the same device in (6.2) we run into the difficulty that the additional appropriations may be such that over an infinite period we will obtain a return of an infinite number of man-hours. To avoid this common difficulty, the usual approach is to introduce a discount factor. One man-hour next year is worth only α, $0 < \alpha < 1$, man-hours now.

There are various rationalizations for the introduction of this factor, depreciation, inflation, uncertainty, etc. It is perhaps most honest to think of it as a mathematical scheme for the elimination of infinities.

Using this factor, (6.2) becomes

$$f_k(x) = \max_{\substack{y_1 + y_2 = x \\ y_1,\ y_2 \ge 0}} [g(y_1) + h(y_2) + \alpha f_{k+1}(ay_1+by_2+c(x-y_1-y_2) + r_k)]. \tag{7.4}$$

We shall encounter the discount factor again in the optimal inventory problem.

Although the introduction of the infinite-stage return simplifies the problem in some ways, it complicates it in others. Attendant with the entrance of infinite processes are all the difficulties of existence of a solution and uniqueness of a solution.

To illustrate this consider (7.2). It was derived from (4.4) by letting $N\to\infty$. One solution will be $\lim_{N\to\infty} f_N(x)$, provided that it exists. However, there is no longer a unique solution since the addition of a constant to any solution will yield a new solution. We must tie the solution down by specifying that $f(0) = 0$.

It is now not difficult to demonstrate, cf. [6], under simple sensible conditions on g and h that there is a unique bounded solution to (7.2) which is 0 at 0. This result is not merely mathematical embonpoint, but an actual necessity if we wish to be sure that the equation describes

the actual process. Just as in dealing with algebraic equations we must avoid extraneous roots, so in dealing with these functional equations we must avoid extraneous solutions.

Various other existence and uniqueness theorems will be found in [6], [8].

8. ANALYTIC RESULTS

Let us restrict ourselves to a consideration of Version 1. The results are typical of what may be obtained by the use of relatively simple methods.

For proofs of the results cited below we refer to [6], [8].

THEOREM 1. If g(x) and h(x) are both strictly convex functions of x, an optimal policy requires that y equal 0 or x.

This result holds for the solution of (7.2), and for the solution of (4.4).

The situation where g and h are both concave is more complicated.

THEOREM 2. Let

$$g(0) = h(0) = 0, \tag{8.1}$$

$$g'(x),\ h'(x) \geq 0 \text{ for } x \geq 0, \tag{8.2}$$

$$g''(x),\ h''(x) \leq 0 \text{ for } x \geq 0, \tag{8.3}$$

and consider the sequence of equations

$$f_1(x) = \max_{0 \leq y \leq x} [g(y) + h(x-y)] \tag{8.4}$$

$$f_{n+1}(x) = \max_{0 \leq y \geq x} [g(y) + h(x-y) + f_n(ay+b(x-y))],\ n = 1, 2, \ldots$$

For each n, there is a unique $y_n = y_n(x)$ which yields the maximum. If $b \leq a$, we have $y_1 \leq y_2 \leq y_3 \leq \ldots$, and the reverse inequalities for $b \geq a$. In particular, if $y_n(x) = x$, for some n in the case $b \leq a$, then $y_m(x) = x$ for $m \geq n$.

This result is useful for approximation purposes since y_1, y_2 and even y_3 may be determined by hand-computation quite quickly.

Even when g and h are convex and we know that y = 0 or x, it is not easy to determine which is the correct y-value. The following result is useful for approximation purposes.

THEOREM 3. The solution of

$$F(x) = \max_{0 \leq y \leq x} [cx^d + F(ax),\ ex^f + F(bx)] \tag{8.5}$$

is given by, in the case where $f > d > 1$

$$(8.6) \qquad \begin{aligned} y &= x \quad \text{for } 0 \le x \le x_0 \\ &= 0 \quad \text{for } x_0 \le x, \end{aligned}$$

where

$$(8.7) \qquad x_0 = [(c/(1-a^d))/(e/(1-b^d))]^{1/(f-d)}.$$

Another particular case where the solution may be obtained simply is that where g and h are quadratic in x.

9. COMPUTATIONAL TECHNIQUES

In the construction of mathematical models of physical problems of all types, there is a perpetual conflict between realistic description and analytic simplicity. It is consequently to be expected that the more faithful the transcription of reality, the smaller will be the probability of being able to handle the problem by analytic means alone.

It follows that the mathematical approach aims at two goals:

(9.1)

1. To describe a uniform technique for solving problems of small dimension.
2. To furnish a sufficiently broad basis for the approximate solution of problems of large dimension, with means for increasing the accuracy of the approximation at the expense of additional cost in time and effort.

To cite an example in a different field, consider the problem of determining the trajectory of a shell. We know that we must consider the aerodynamic characteristics of the shell, the air resistance as a function of velocity, the velocity of the wind, the curvature of the earth, the change in muzzle characteristics as a function of the number of previous firings, and a multitude of other factors. Nevertheless, one starts the problem by considering the parabolic trajectory of a point particle in a vacuum subject only to the force of gravity.

The moral of the tale is that the whole purpose of examining simple models possessing certain properties of the original is to obtain a vantage point from which to proceed to the exploration of more and more complicated models.

In the study of these dynamic programming problems, we possess an approximation technique which is not available for the study of most of the functional equations of analysis. This tool is the duality that exists between the solution of the functional equation and the optimal policy.

Referring to (7.2), we see that given f(x) we can determine a maximizing y, and indeed all maximizing y. Conversely, if we know y as a function of x, we can obtain f(x) by straightforward iteration. Thus the function determines all optimal policies, and an optimal policy determines the function.

We see then that in attempting to find an optimal policy, which is our main goal, we have our choice of guessing an initial approximation in function space or in policy space.

In most cases, it will be far preferable to guess an initial policy, since experience will have provided a firm basis for a tentative solution. This is essentially the situation in Poker involving three or more people. Although there is no mathematical theory[1] to dictate correct play the experience of years of play furnishes a heuristic set of canons of good play which are rather costly to disobey.

It is precisely in this guess of an initial approximation, to be used as the first of sequence of successive approximations, that actual experience plus the theoretical insight gained from the consideration of simplified models may be combined most fruitfully.

Furthermore, a tremendous advantage possessed by the method of approximating by means of a policy, rather than a function, lies in the fact that successive approximations will automatically yield better results.

Let us amplify these remarks, considering (7.2). If an initial guess is $f = f_g$, we obtain approximations by considering

$$f_{N+1}(x) = \max_{0 \le y \le x} [g(y) + h(x-y) + f_N(ay + b(x-y))]. \tag{9.2}$$

Under the same conditions that guaranteed existence and uniqueness, it may be demonstrated that $f_N(x)$ converges to f. However, the convergence is not necessarily monotone.

If we take the initial guess to be f_1 as in (4.2), the convergence will be monotone, but usually too slow. If, however, the functions $f_N(x)$ are also of interest, this method has the merit of providing a great deal of additional information.

Let us now consider the technique of approximating by means of policies. We have a number of choices, for example

(9.3)
(a) We may choose at each step the y which maximizes $g(y)+h(x-y)$.
(b) We may approximate to $g(x)$ and $h(x)$ by functions of the form ax^d, or by means of quadratics, and use the y's obtained from these solutions.
(c) We may use the unit cost as a criterion and choose y so that

$$\frac{g(y)}{(1-a)y} = \frac{h(x-y)}{(1-b)(x-y)}.$$

In either case, we have a means for determining an approximate expression for y as a function of x, Using this approximate $y = y_G$ we compute f(x) recurrently and call the result F_g.

This function F_g satisfies the equation

$$F_g(x) = g(y_G) + h(x-y_G) + F_g(ay_G+b(x-y_G)), \quad (y \equiv y_G(x)), \tag{9.4}$$

and thus may be computed recurrently.

Consider the second approximation,

$$f_2(x) = \max_{0 \le y \le x} [g(y) + h(x-y) + F_g(ay+b(x-y))]. \tag{9.5}$$

[1] Not, however, for two person Poker! cf [3], [11].

It follows that for all y in [0, x] we have

$$f_2(x) \geq g(y) + h(x-y) + F_g(ay+b(x-y)), \tag{9.6}$$

and thus

$$f_2(x) \geq F_g(x) = g(y_G) + h(x-y_G) + F_g(ay_G + b(x-y_G)). \tag{9.7}$$

Since

$$f_3 = \operatorname*{Max}_{0 \leq y \leq x} [g(y) + h(x-y) + f_2(ay + b(x-y))], \tag{9.8}$$

it follows by comparison of (9.5) and (9.8) that $f_3 \geq f_2$ since $f_2 \geq F_g$. Consequently, it follows inductively that $f_{N+1} \geq f_N \geq \ldots \geq f_G$, with strict inequalities unless $F_g = f$.

10. UNCERTAINTY

We have, in the previous sections, assumed that the outcome of each expenditure was perfectly definite, resulting in a certain number of man-hours and permitting a fixed salvage value. Actually, of course, we can predict a mean number of man-hours to be obtained from an expenditure of y dollars on A-equipment, and we can make a guess as to the salvage value, depending upon whether the economy is inflating or deflating and depending upon the state of the equipment at the end of the year.

How do we make definite statements as to the form of an optimal policy in the face of this uncertainty?

It is first of all clear that we cannot maximize the total man-hours obtained over the N-year period, since this quantity is an uncertain number, depending on a multitude of factors, many of which are completely beyond our control. Consequently we must agree to use some average value. The simplest such average is the average number of man-hours obtained over the N-year period; in mathematical parlance, the expected total man-hours.

Another quantity of importance is the probability that the total man-hours exceed M, which is to say the probability that we obtain at least M man-hours altogether. In many cases, this is a safer criterion than the expected value.

Here we shall consider only the simplest case, the average or expected value.

Let us assume, for simplicity, that if we spend y dollars to purchase A-equipment, there are only two possibilities:

(10.1) (a) With probability p_1 we will obtain $g_1(y)$ man-hours and have a salvage value $a_1 y$.
(b) With probability $p_2 = 1 - p_1$ we will obtain $g_2(y)$ man-hours and have a salvage value $a_2 y$.

Similarly, if (x-y) dollars are spent for B-equipment,

(10.2) (a) With probability q_1 we will obtain $h_1(x-y)$ man-hours and have a salvage value $b_1(x-y)$.

(b) With probability $q_2 = 1 - q_1$ we will obtain $h_2(x-y)$ man-hours and have a salvage value $b_2(x-y)$.

Here $0 < a_1, a_2, b_1, b_2 < 1$. If we define, for an infinite stage process,

(10.3) $f(x)$ = expected number of total man-hours obtained starting with x dollars and using an optimal policy,

we see, by means of a simple enumeration of possibilities, as before, that $f(x)$ satisfies the equation

$$f(x) = \max_{0 \le y \le x} [p_1q_1(g_1(y) + h_1(x-y) + f(a_1y + b_1(x-y)) + p_1q_2(g_1(y) + h_2(x-y) + f(a_1y + b_2(x-y))) + p_2q_1(g_2(y) + h_1(x-y) + f(a_2y + b_1(x-y))) + p_2q_2(g_2(y) + h_2(x-y) + f(a_2y + b_2(x-y)))] \tag{10.4}$$

This equation is essentially of the same type as those obtained above in the deterministic case. We see that once we have introduced the concept of expected value we are essentially dealing with a deterministic process.

11. OPTIMAL INVENTORY

A problem from quite a different source which leads to closely related functional equations is that of optimal inventory. Essentially, the problem is the following. We have a quantity of equipment on hand and have the prerogative of ordering more at a certain cost per unit ordered. At various times, requests will come in for various quantities of this equipment. Unfortunately, the precise quantity that will be requested at any future time is not known. Usually, we have some idea of the levels of demand and the probability that any particular quantity will be requested.

If we order excessively to ensure that we can meet any conceivable demand, we are faced with problems of obsolescence, of cost of storage, and with the fact that the money could be more profitably used elsewhere.

If we keep the supply level too low, we are faced with the unpleasant consequences of a shortage of material, and it is difficult to assess a proper penalty for shortage. In economic situations it may mean that we must make up the shortage with more costly equipment, or at a greater cost in time, which is essentially the same thing.

A mathematical discussion of problems of this type was first given by Arrow, Harris and Marschak, [2], and then together many extensions by Dvoretzky, Wolfowitz and Kiefer, [13].

Let us here formulate a simple problem which will illustrate how similar these problems are mathematically to those discussed in the preceding sections.

Let us assume that orders have to be made at the beginning of each year, and that the requests occur solely in the middle of the year, and that only one order and one request per year

are made. Let us consider only two types of requests, large and small, in quantity ℓ and s respectively, and assume a probability p_1 of a large order and p_2 of a small order.

The cost of ordering a quantity y is taken to be g(y), with g(0) = 0.

In order to simplify the problem still further, let us assume that the process is repeated year after year without end, and that the cost is discounted each year by an amount a, the usual device to insure finiteness of cost.

Finally let us agree that if a request can be filled, it will be filled, while if it exceeds the stock the request will be filled as far as capable and a penalty proportional to shortage will be levied. If the shortage is z, the penalty will be h(z).

Let us define

(11.1) $f(x)$ = expected total cost over the infinite period given an initial stock of x, using an optimal policy.

Suppose that y is ordered at the beginning of the first year. The expected cost will be compounded of two parts. The first is the immediate cost due to ordering. The second will be the expected cost over the future. This cost will be

(11.2)
$$a[p_1 f(x+z-\ell) + p_2 f(x+z-s)], \text{ for } x + z \geq \ell,$$

$$a[p_1 h(\ell-x-z) + p_1 f(0) + p_2 + (x+z-s)], \text{ for } \ell \geq x + z \geq s,$$

$$a[p_1 h(\ell-x-z) + p_2 h(s-x-z) + (p_1+p_2) f(0)], \text{ for } \ell \geq x + z.$$

Let us call this quantity aF(x+z). Then the total expected cost will be g(z) + aF(x+z). Since z is to be chosen to minimize the total expected cost, we obtain the functional equation

(11.3)
$$f(x) = \operatorname{Min}_{z \geq 0} [g(z) + aF(x+z)].$$

The approximation methods discussed in Section 8 may now be used to obtain a numerical solution. There are also various simple cases in which exact solutions may be obtained, cf. the above references and [6].

12. A SMOOTHING PROBLEM

In the previous sections we have illustrated the use of the functional equation approach by means of simple examples. Let us now discuss briefly the application of another very important tool, the method of continuous approximations.

As an example, let us consider a typical "smoothing" problem. For the sake of simplicity, let us make it determinate. Suppose that we are given a work schedule over the time period $0 \leq t \leq T$, which tells us that at t = k we will require r(k) employees to perform the job assigned at that time. Let t take only the values 0, 1, 2, ..., T.

As is often the case, r(k) will be an oscillating function, sometimes increasing, sometimes decreasing. Whenever we have more than r(k) employees, we are losing the salaries of the unused employees. On the other hand, for many reasons it is expensive to hire new, possibly untrained, employees.

The problem is to determine how many employees we should have at each time so as to balance the loss due to overstaffing against the loss due to rapid buildup.

Let x_k be the number of employees at time k, assuming that $x_0 = r_0$. Then a first cost will be $\beta(x_k - r_k)$, due to overstaffing. A second cost will be the cost involved in bringing x_k up to a possibly higher level at the next time k+1. This cost will be taken to be

$$\alpha(x_{k+1} - x_k) \quad \text{if} \quad x_{k+1} - x_k > 0 \tag{12.1}$$

$$0 \quad \text{if} \quad x_{k+1} - x_k \leq 0,$$

where k and β are positive constants. These two expressions can be combined into one, namely $\alpha \max(x_{k+1} - x_k, 0)$. The total cost over the entire time period will be

$$J_N(x) = \sum_{k=0}^{N-1} [\beta(x_k - r_k) + \alpha \max(x_{k+1} - x_k, 0)] + \beta(x_N - r_N). \tag{12.2}$$

We now wish to choose the sequence x_k, subject to the restriction $x_k \geq r_k$, so as to maximize $J_N(x)$.

In order to obtain an approximate solution to this problem, we consider the continuous analogue to (2), namely

$$J_T(x) = \int_0^T [\beta(x(t) - r(t)) + \alpha \max(\frac{dx}{dt}, 0)]\, dt. \tag{12.3}$$

We wish to determine the x(t) which minimize $J_T(x)$. The constraint is $x(t) \geq r(t)$.
This problem may be solved completely. For the details we refer to [10].
For further illustrations of the use of continuous analogues, see [5], [9], and [12].

REFERENCES

[1] Alchian, A., "Economic Replacement Policy," Rand Report 224, April, 1952.

[2] Arrow, K. D., Harris, T. E., and Marschak, J., "Optimal Inventory Policy," Econometrica, Vol. 19, No. 3, (July, 1951), pp. 250-272.

[3] Bellman, R., "On Games Involving Bluffing," Rendiconti del Palermo, (2), Vol. 1 (1952), pp. 1-18.

[4] Bellman, R., "An Optimal Inventory Equation," Rand Paper P-432.

[5] Bellman, R., "Bottleneck Problems and Dynamic Programming," Proc. Nat. Acad. Sci., Vol. 39 (1953), pp. 947-951.

[6] Bellman, R., "A Problem in the Theory of Dynamic Programming," Econometrica, to appear.

[7] Bellman, R., "On the Theory of Dynamic Programming," Proc. Nat. Acad. Sci., Vol. 38 (1952), pp. 716-719.

[8] Bellman, R., "Some Functional Equations in the Theory of Dynamic Programming," Proc. Nat. Acad. Sci., Vol. 39 (1953), pp.

[9] Bellman, R., Glicksberg, I., and Gross, O., "Some Variational Problems in the Theory of Dynamic Programming," Proc. Nat. Acad. Sci., Vol. 39 (1953), pp. 298-301.

[10] Bellman, R., "Smoothing Problems - 1," Rand unpublished paper.

[11] Bellman, R., and Blackwell, D., "Some Two-person Games Involving Bluffing," Proc. Nat. Acad. Sci., Vol. 35 (1949), pp. 600-605.

[12] Bellman, R., and Lehman, S., "On the Continuous Gold-Mining Equation," Proc. Nat. Acad. Sci., to appear.

[13] Dvoretzky, A. J., Kiefer, J., and Wolfowitz, J., "The Inventory Problem—I: Case of Known Distributions of Demand" and "The Inventory Problem—II: Case of Unknown Distributions of Demand," Econometrica, Vol. 20 (1952), pp. 187-222.

* * *

THE CATERER PROBLEM

Walter Jacobs
U. S. Air Force

The Caterer Problem is a paraphrased version of a practical military problem which arose in connection with the estimation of aircraft spare engine requirements. In this paper the problem is put into the form of a linear programming problem, and by several transformations an explicit solution of that linear programming problem is obtained for a special case. The technique used may be applicable to certain other linear programming problems.

1. INTRODUCTION

A number of general methods are available for computing solutions to linear programming problems—that is, for finding a solution to a system of linear inequalities which maximizes or minimizes a linear function of the variables. There are direct procedures such as the Simplex Method [1], [2], and the Elimination Method [3], as well as iterative techniques [1], [3] (numbers in square brackets refer to the bibliography).

As the standard form of linear programming problem it will be convenient to deal with sets of linear equations in non-negative unknowns, rather than with linear inequalities. A problem can always be converted from one to the other of these forms. Thus we will take as the general linear programming problem the system

$$
\begin{aligned}
& a_{11}\ x_1 + a_{12}\ x_2 + \ldots + a_{1n}\ x_n = b_1 \\
& a_{21}\ x_1 + a_{22}\ x_2 + \ldots + a_{2n}\ x_n = b_2 \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& a_{m1}\ x_1 + a_{m2}\ x_2 + \ldots + a_{mn}\ x_n = b_m \qquad (1.1) \\
& x_1 \geq 0;\ x_2 \geq 0;\ \ldots;\ x_n \geq 0 \\
& c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \ \text{a minimum}
\end{aligned}
$$

To solve such a problem by the Simplex method requires somewhat more computation than to solve by elimination a set of m linear equations in m unknowns. Consequently, even for problems of moderate size, say 25 equations in 50 variables, a digital computer may be necessary. Experience thus far with other general procedures implies that they are likely to be slower than the Simplex method.

When linear programming is applied to dynamic problems, in which a linear system of activities is studied over a succession of discrete time periods, the number of equations can easily become large, yielding a problem which is extremely lengthy to handle by the Simplex method. On the other hand, such problems often will have coefficient matrices of special form, with recurring sub-matrices and very many zero elements. The question then

arises whether it is possible to take advantage of the special nature of the coefficient matrix and thus obtain a solution which requires less work than the general procedures.

Special solutions of this kind have been worked out for a number of problems, non-dynamic as well as dynamic; a well-known example is the Hitchcock-Koopmans Transportation problem (articles on this are contained in references [1] and [3]). The problem discussed here is amenable to such a solution—in fact in its simplest form it can be solved as an explicit function of the data. The few prior instances where explicit solutions have been found for linear programming problems have involved much simpler systems.

Methods used in a problem such as this are of general interest to the extent that they are applicable to other problems. The feature of the solution to the Caterer problem which seems most promising for wider application is the use of a transformation to simplify the initial problem. This requires a brief amplification.

It is ordinarily not possible, when nothing is known about the solution, to simplify a set of linear equations in non-negative variables (or generally, a set of linear inequalities) by making a linear transformation. As a particular case of this, elimination of variables does not help. To see why this is so, consider the set (1.1) and suppose $a_{11} \neq 0$; solve the first equation for x_1, and eliminate x_1 from the remaining equations. This procedure has not, however, reduced the problem to a simpler set of m-1 equations in n-1 variables, for we must retain the further condition on the variables $x_2, x_3, \ldots, x_n$ that

$$\frac{1}{a_{11}} [b_1 - a_{12} x_2 - a_{13} x_3 - \ldots - a_{1n} x_n] \geq 0, \tag{1.2}$$

corresponding to the requirement that x_1 be non-negative.[1]

The presence of a linear function to be minimized (or maximized) may alter the situation. As the Caterer problem will illustrate, it is sometimes possible to find a transformation which converts a linear programming problem into a simpler problem related to the original in the following way: the optimal solutions of the transformed problem always corresponds to optimal solutions of the original problem, even though there are feasible solutions of the transformed problem corresponding to which there are no feasible solutions of the initial system.

2. STATEMENT OF THE PROBLEM: A caterer knows that in connection with the meals he has arranged to serve during the next n days, he will need $r_j (\geq 0)$ fresh napkins on the j^{th} day, $j = 1, 2, \ldots, n$. Laundering normally takes p days; that is, a soiled napkin sent for laundering immediately after use on the j^{th} day is returned in time to be used again on the $(j + p)^{th}$ day. However, the laundry also has a higher-cost service which returns the napkins in $q < p$ days (p and q integers). Having no usable napkins on hand or in the laundry, the caterer will meet his early needs by purchasing napkins at a cents each. Laundering costs b and c cents a napkins for the normal and high-cost service respectively. How does he arrange matters to meet his needs and minimize his outlays for the n days?

[1] Of course, if $a_{11} \geq 0$, and $a_{12} \leq 0, a_{13} \leq 0, \ldots, a_{1n} \leq 0$ and $b_1 \geq 0$, the condition is redundant, since it is implied by $x_2 \geq 0, \ldots, x_n \geq 0$.

In the actual military application to aircraft spare engine requirements, "laundering" corresponds to overhaul of engines removed from the aircraft because of failure, and the unit of time is a month rather than a day. As long as the time periods covered in the computation extend sufficiently beyond the period of peak operation of the engines, the use of a formulation which ignores the effect of requirements after n periods is acceptable. A more general formulation, which would apply to other economic situations involving the choice between more rapid and cheaper methods of repairing spare parts, would have to consider starting and ending inventories and other complications as well. Some consideration has been given to the more general problem, but it is not the concern of the present paper.

Before expressing the problem algebraically, two conventions of notation will be stated. The subscript j throughout has the range 1, 2, . . . , n; every equation involving j is to hold for the entire range of values. Where this results in the appearance of quantities with subscripts outside the range, such quantities are always zero. The cumulative of a quantity designated by a small letter is represented by the corresponding capital letter; e.g.,

$$X_j = \sum_{i=1}^{j} x_i \quad ; \quad R_{j+p} = \sum_{i=1}^{j+p} r_i . \tag{2.1}$$

Let x_j represent the napkins purchased for use on the j^{th} day; the remaining requirements, if any, are supplied by laundered napkins. Of the r_j napkins which have been used on that day plus any other soiled napkins on hand, let y_j be the number sent for laundering under normal service and z_j the number under the rapid service. Finally, since soiled napkins need not be sent to the laundry immediately after use, let s_j be the stock of soiled napkins on hand after $y_j + z_j$ have been shipped to the laundry; they will be available for laundering on the next day together with the r_{j+1} used that day. Consequently,

$$y_j + z_j + s_j - s_{j-1} = r_j . \tag{2.2}$$

That is, the napkins used on the j^{th} day are equal to the number sent to the laundry plus the change in the stock of soiled napkins.[2]

The returns from the laundry are equal to the amounts shipped for normal service p days earlier plus the amounts shipped for rapid service q days earlier. Together with purchases, these provide for the needs on the same day.[3]

$$x_j + y_{j-p} + z_{j-q} = r_j . \tag{2.3}$$

The total cost to be minimized, subject to these constraints, is

$$\sum_j (ax_j + by_j + cz_j)$$

[2]Note that, when j = 1, the quantity s_0 occurs, and by the convention mentioned earlier, it is taken as zero.

[3]The set (2.3) does not provide for stocks of fresh napkins; it is easy to show that if the problem is changed to allow such stocks, the solutions are equivalent in cost to those of the present set, since the additional possibilities merely involve advance purchases or laundering.

or, using the convention exemplified in (2.1),

$$aX_n + bY_n + cZ_n .$$

Here $a > b$, $c > b$.

It is clear that, for a minimum-cost solution, napkins are not laundered unless they can be returned by the n^{th} day. That is

$$y_{n-p+1} = y_{n-p+2} = \cdots = y_n = 0$$

$$z_{n-q+1} = z_{n-q+2} = \cdots = z_n = 0 .$$

These values of y_j, z_j are accordingly to be omitted hereafter wherever they are implied. In particular, the cost-function becomes

$$aX_n + bY_{n-p} + cZ_{n-q} . \tag{2.4}$$

Collecting the relations (2.2), (2.3), (2.4), the Caterer problem is given by the system:

$$\begin{aligned} &x_j + y_{j-p} + z_{j-q} = r_j \\ &y_j + z_j + s_j - s_{j-1} = r_j \\ &x_j \geq 0;\ y_j \geq 0;\ z_j \geq 0;\ s_j \geq 0 \\ &aX_n + bY_{n-p} + cZ_{n-q} \text{ a minimum.} \end{aligned} \tag{2.5}$$

The principal result of the paper is the explicit solution of the system (2.5) in the special case $q = p-1$; this is given by the theorem below. However, in proving the theorem, the system is substantially simplified, and in this reduced form it is then solvable by a special procedure which has been described elsewhere [4].

We first define the symbol

$$f^+ = \max(f, 0) .$$

Also we define $f^{(m)}$ to be the m^{th} largest of the set of non-negative quantities f_j; if $m > n$, $f^{(m)} = 0$.

THEOREM: A solution to the system

$$\begin{aligned} &x_j + y_{j-p} + z_{j-p+1} = r_j \\ &y_j + z_j + s_j - s_{j-1} = r_j \\ &x_j \geq 0;\ y_j \geq 0;\ z_j \geq 0;\ s_j \geq 0 \end{aligned} \tag{2.6}$$

which minimizes $aX_n + bY_{n-p} + cZ_{n-p+1}$ is given by

$$
\begin{aligned}
z_j &= (H_{j+p-1} - H^{(m)})^+ \\
x_j &= \max_{i \le j} (R_i - R_{i-p} - z_{i-p+1}) - \max_{i < j} (R_i - R_{i-p} - z_{i-p+1}) \\
y_j &= r_{j+p} - x_{j+p} - z_{j+1} \\
s_j &= R_j - Y_j - Z_j ,
\end{aligned}
\tag{2.7}
$$

where

$$H_j = (R_j - R_{j-p} - \max_j (R_j - R_{j-p+1}))^+ ,$$

and m is the least integer greater than $\frac{a-b}{c-b}$.

The proof of the theorem is long, and it is convenient to divide it into several steps which are summarized below. All but the last two of these steps apply to the case of general q, i.e., the system (2.5), and they are stated accordingly:

1. It is shown that the cost function in (2.5) may be put into a more convenient form.
2. (Lemma 1). With this form of cost function, the original problem is simplified by a transformation.
3. (Lemma 2). A number of the resulting inequalities are found to be redundant.
4. When q = p-1, a further set of inequalities are redundant.
5. The minimal solution to the remaining set, given by Lemma 3, corresponds to the solution (2.7) of the theorem.

The first two of these steps are covered in the following section.

3. THE REDUCED PROBLEM

To put the cost function into a more convenient form, we sum the set (2.3) for all j. This gives, with the notation described in (2.1)

$$X_n + Y_{n-p} + Z_{n-q} = R_n . \tag{3.1}$$

Since we may change the cost function by an additive or positive multiplicative constant without affecting the solution, we subtract b times the left-hand side of (3.1) from (2.4) and divide by c-b > 0; the new function to be minimized becomes

$$dX_n + Z_{n-q} , \qquad d = \frac{a-b}{c-b} .$$

In the following lemma, we have replaced X_n by w in (3.2) to emphasize the disappearance of the individual x_j :

LEMMA 1: To any solution of the system in the non-negative variables w, z_j, u_j, v_j :

$$w + Z_{j-q} - Z_{j-p} - u_j = R_j - R_{j-p}$$

$$z_{j-p} + v_j = r_j \tag{3.2}$$

$$dw + Z_{n-q} \text{ a minimum}$$

corresponds a solution of the system in the non-negative variables x_j, y_j, z_j, s_j

$$x_j + y_{j-p} + z_{j-q} = r_j$$

$$y_j + z_j + s_j - s_{j-1} = r_j \tag{3.3}$$

$$dX_n + Z_{n-q} \text{ a minimum}$$

with the same values of the z_j and with $X_n = w$.

The system (3.2) is obtained from (3.3) by transforming the variables in the following way: Sum the first equation in (3.3) from $j = 1$ to $j = j$, and the second equation from 1 to $j - p$. There results

$$X_j + Y_{j-p} + Z_{j-q} = R_j$$

$$Y_{j-p} + Z_{j-p} + s_{j-p} = R_{j-p} ,$$

and eliminating Y_{j-p} between these equations gives

$$X_j + Z_{j-q} - Z_{j-p} - s_{j-p} = R_j - R_{j-p} . \tag{3.4}$$

Writing

$$u_j = X_n - X_j + s_{j-p} \tag{3.5}$$

and substituting for $X_j - s_{j-p}$ in (3.4), we obtain the first equation (3.2) with $X_n = w$. The second equation in (3.2) follows from the first equation in (3.3) with

$$v_j = x_j + y_{j-p} \tag{3.6}$$

From the fact that the u_j and v_j defined in (3.5) and (3.6) are linear combinations with non-negative coefficients of the x_j and s_j or y_j, it follows that from any non-negative solution of the equations in (3.3), we can derive a non-negative solution of the equations in (3.2) by the above transformation. Thus the feasible solutions of (3.3) yield corresponding feasible solutions of (3.2).[4] The converse is not true: there are feasible solutions to (3.2) which cannot be obtained by transforming a solution of (3.3) in this way.

As will be shown, however, feasible solutions of (3.3) always exist corresponding (under the given transformation) to optimal solutions of (3.2). It follows at once from the equality of the cost functions in the two problems that these feasible solutions are optimal for (3.3), since otherwise the optimal solution for (3.3) would have a lower cost, and there would be a feasible solution for (3.2) with this lower cost.

Consequently, to complete the proof of the lemma, it must be shown that given the u_j, v_j of an optimal solution to (3.2), non-negative values of x_j, y_j, s_j satisfying (3.5) and (3.6) can be determined, or equivalently, that there exist non-negative values of x_j satisfying

$$X_n = w$$

$$X_n - X_j \leq u_j \tag{3.7}$$

$$x_j \leq v_j$$

where u_j and v_j are optimal solutions in (3.2).[5] In fact, choose

$$X_j = w - \min_{i \leq j} u_i \; ; \tag{3.8}$$

X_j thus defined is monotonic and non-decreasing, hence x_j is non-negative. For an optimal solution to (3.2), $\min u_j = 0$; otherwise w and every u_j could be decreased by a fixed amount. The first two equations in (3.7) follow from this. The third equation holds for every x_j which vanishes, and the proof of the lemma will be complete if we show that the result is true also for positive x_j.

Let k be a value of j for which $x_k > 0$. By (3.6) it is sufficient to show that $y_{k-p} \geq 0$, and clearly we may suppose that $k > p$. Note that $x_k > 0$ implies in (3.8) that

$$X_k = w - u_k = X_n - u_k$$

and this with (3.5) shows that $s_{k-p} = 0$. Moreover, from $x_k > 0$ and (3.8) it follows that $u_i > 0$ when $i < k$. But then $z_{k-p} = 0$; otherwise, z_{k-p} and each of $u_{k-p+q}, u_{k-p+q+1}, \cdots, u_{k-1}$ all could be reduced equally without otherwise affecting (3.2), contradicting the optimality of the solution.

[4] In the usual terminology, a feasible solution is a set of non-negative values satisfying the constraints. An optimal solution is a feasible solution which minimizes the cost function.

[5] This guarantees the existence of non-negative solutions to the system consisting of (3.4) together with the second set of equations in (3.3). But this is equivalent to the existence of a feasible solution to (3.3) since the transformation yielding (3.4) is non-singular.

Finally, using $z_{k-p} = s_{k-p} = 0$ in the second equation of (3.3),

$$y_{k-p} = r_{k-p} + s_{k-p-1} \geq 0 ,$$

which completes the demonstration of the lemma.

A number of remarks about the lemma may be made in passing. In the first place, it still holds if the cost function in (3.2) is replaced by

$$dw + \Sigma d_j z_j , \qquad d_j > 0 , \tag{3.9}$$

with a corresponding change in (3.3). Secondly, as the proof makes clear, the conditions

$$y_j \geq 0, \qquad s_j \geq 0$$

of (3.3) can be relaxed to

$$y_j + x_{j+p} \geq 0$$

$$s_j + \sum_{j+p+1}^{n} x_i \geq 0$$

without affecting the set of solutions. Lastly, a restatement of the constraints in (3.2) and (3.3) as inequalities is of some interest:

$$w + Z_{j-q} - Z_{j-p} \geq R_j - R_{j-p}$$

$$r_j \geq z_{j-q} \geq 0 \tag{3.2'}$$

$$w \geq 0$$

and

$$x_j + y_{j-p} + s_{j-q} - r_j$$

$$Y_j + Z_j \leq R_j \tag{3.3'}$$

$$x_j \geq 0; \quad y_j \geq 0; \quad z_j \geq 0 .$$

Thus, the lemma might have stated the correspondence between the solutions of (3.2') and (3.3') which minimize (3.9).

4. THE SOLUTION OF THE REDUCED PROBLEM

Eliminating the z_j from the first inequality (3.2') by using the second inequality, it is easily seen that

$$w \geq R_j - R_{j-q}$$

We set

$$\bar{w} = w - \max(R_j - R_{j-q}) ; \tag{4.1}$$

then $\bar{w} \geq 0$. Rewrite the first equation (3.2') as

$$\bar{w} + Z_{j-q} - Z_{j-p} \geq R_j - R_{j-p} - \max(R_j - R_{j-q}) ;$$

since the left-hand side is non-negative, this may be strengthened by replacing the right-hand side by zero when it is negative:

$$\begin{aligned} &\bar{w} + Z_{j-q} - Z_{j-p} \geq H_j \\ &H_j = (R_j - R_{j-p} - \max(R_j - R_{j-q}))^+ . \end{aligned} \tag{4.2}$$

But $H_i = 0$, $(i = 1, 2, \ldots, q)$, and the first q equations in (4.2) are redundant; thus, we have proved

LEMMA 2: The system (3.2) is equivalent to

$$\begin{aligned} &\bar{w} + Z_k - Z_{k-p+q} \geq H_{k+q} \\ &z_k \leq r_{k+q} \\ &z_k \geq 0; \quad \bar{w} \geq 0 \\ &d\bar{w} + Z_{n-q} \text{ a minimum,} \end{aligned} \tag{4.3}$$

where k has the range $(1, 2, \ldots, n - q)$.

It does not appear that a useful explicit solution to this system can be found for general values of q.[6] In the remainder of this section we consider the case $q = p - 1$, and conclude the proof of our original theorem. For this case, (4.3) reduces to

$$\begin{aligned} &\bar{w} + z_k \geq H_{k+q} \\ &z_k \leq r_{k+q} \\ &z_k \geq 0; \quad \bar{w} > 0 \\ &d\bar{w} + Z_{n-q} \text{ a minimum} \end{aligned} \tag{4.4}$$

[6]However, a straightforward procedure for computing solutions is given in [4].

We note that if $H_j = 0$, $H_j \le r_j$; if $H_j > 0$,

$$\begin{aligned} H_j &= R_j - R_{j-q-1} - \max(R_j - R_{j-q}) \\ &\le R_j - R_{j-q-1} - (R_{j-1} - R_{j-q-1}) \\ &\le r_j . \end{aligned}$$

It follows that the conditions $z_k \le r_{k+q}$ in (4.4) are redundant. For, in the system

$$\begin{aligned} &\bar{w} + z_k \ge H_{k+q} \\ &s_k \ge 0; \quad \bar{w} \ge 0 \\ &d\bar{w} + Z_{n-q} \text{ a minimum,} \end{aligned} \tag{4.5}$$

any minimal solution must evidently satisfy

$$z_k \le H_{k+q} \le r_{k+q} ;$$

if any z_k exceeds the corresponding H_{k+q}, that z_k can be reduced without changing the other unknowns or violating the conditions of the problem. Thus, (4.5) implies (4.4).[7]

The solution to (4.5) is implied by

LEMMA 3: Let $a_1 \ge a_2 \ge \ldots \ge a_n \ge 0$. A solution to

$$\begin{aligned} &w + z_j \ge a_j \\ &w \ge 0; \quad z_j \ge 0 \end{aligned} \tag{4.6}$$

which minimizes $dw + Z_n$ is given by

$$w = a_m, \quad z_j = (a_j - a_m)^+$$

where m is the least integer greater than d. If $d > n$, the minimum solution is $w = 0$, $z_j = a_j$.

PROOF: For any value of w, Z_n is a minimum when

$$z_j = (a_j - w)^+ .$$

---

[7] The corresponding simplification is not possible when $p - q > 1$, since the most that we can conclude by the analogous argument is
$Z_k - Z_{k-p+q} \le R_{k+q} - R_{k-p+2q}$, which does not imply $z_k \le r_{k+q}$ for all k.

With these values of z_j, and with $a_m \le w \le a_{m-1}$,

$$dw + Z_n = dw + \sum_{1}^{m-1} (a_j - w)$$

$$= (d - m + 1)w + \sum_{1}^{m-1} a_j$$

Consequently, for the class of minimal solutions to (4.6), the graph of $(dw + Z_n)$ as a function of w is a polygon, convex upwards, with vertices at $w = a_j$ and slope $(d - m + 1)$ in the interval (a_m, a_{m-1}). Suppose first that d is not an integer, and $d \le n$. At the minimum there must be a negative slope to the left and a positive slope to the right of the corresponding vertex; if this point is a_m, and $a_m < a_{m-1}$,

$$d - m < 0,$$

$$d - m + 1 > 0,$$

and m is the least integer greater than d. (If $d > n$, the graph decreases constantly, and has its minimum at $w = 0$.) If $a_m = a_{m-1}$, the solution is unique, but may be written indifferently a_m or a_{m-1}. Finally, if d is an integer, any w satisfying $a_{d+1} \le w \le a_d$ yields a solution.

We now apply Lemma 3 to (4.5), and obtain

$$\bar{w} = H^{(m)}, \qquad z_k = (H_{k+q} - H^{(m)}) .$$

These values of the z_k may be used in (3.4), which for $q = p - 1$ reduces to

$$X_j + z_{j-p+1} \ge R_j - R_{j-p} .$$

A monotone solution for X_j is obtained by setting

$$X_j = \max_{i \le j} (R_i - R_{i-p} - z_{i-p+1})$$

and this is equivalent to the values in (2.7). By Lemma 1 the corresponding y_j and s_j are non-negative, and the proof of the theorem is complete.

5. CONCLUSION

A graphical interpretation of the solution (2.7) may be of some help in comprehending it. Consider the polygon formed by the broken-line curve with vertices at $(j, R_j - R_{j-p})$. Lop off the peaks with a horizontal line whose ordinate is X_n; it passes through the m^{th} highest vertex unless that vertex has an ordinate less than $\max(R_j - R_{j-p+1})$, which will then be the value of X_n. The distance from the horizontal to the j^{th} vertex, when the vertex is above the horizontal, is z_{j-p+1}; otherwise $z_{j-p+1} = 0$. X_j is the highest vertex on the curve, up to and including the j^{th}, but not exceeding X_n.

The reduction of the system (2.7) to the considerably simpler (4.5) has been justified by an analysis which appears to cast little light on the general conditions under which such a reduction is possible. It is to be expected that a criterion for reduction can be found in terms of the coefficients of the constraining inequalities and the signs of the constant terms and cost coefficients. If this criterion permits the recognition of cases where problems, too large to handle by general procedures, can be brought within the range of practical computing methods, it will materially extend the applications of linear programming to dynamic systems.

In closing, the author wishes to express his appreciation to A. J. Hoffman and E. D. Schell for their helpful suggestions in the preparation of this paper.

REFERENCES

[1] Koopmans, T. C., Ed., Activity Analysis of Production and Allocation, Cowles Commission Monograph 13, John Wiley and Sons, 1951.

[2] Charnes, A., Cooper, W. W., and Henderson, A., An Introduction to Linear Programming, John Wiley and Sons, 1953.

[3] Symposium on Linear Inequalities and Programming, SCOOP Publication No. 10, Headquarters US Air Force (photo-offset).

[4] Gaddum, J. W., Hoffman, A. J., and Sokolowsky, D., "On the Solution to the Caterer Problem," NBS Report 2971, U. S. Dept. of Commerce, National Bureau of Standards (mimeographed).

NEWS AND MEMORANDA

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics.

The Princeton University Logistics Research Project recently issued notes on lectures by W. Fenchel on "Convex Cones, Sets, and Functions" and notes on lectures by H. W. Kuhn on "The Theory of Games." Copies of these lecture notes can be obtained from Professor A. W. Tucker, Fine Hall, Box 708, Princeton University, Princeton, New Jersey.

A seminar on linear programming has been initiated by the National Bureau of Standards under the chairmanship of Dr. Alan J. Hoffman. The meetings are being held in Room 104, Hurst Building of The American University in Washington, D. C. The first meeting was held March 31, 1954 and meetings are being held every two weeks thereafter.

The Fifth Annual Logistics Research Conference is being planned by the Office of Naval Research and The George Washington University Logistics Research Project for the purpose of developing effective plans for the utilization of usage data by the Navy. This conference, to be held at The George Washington University during the week of October 25, 1954, will be a working conference with limited attendance. Only those individuals working in the field or who are in a position to contribute to the conference are being invited.

A meeting of the Office of Naval Research-Bureau of Supplies and Accounts liaison group for monitoring the collection and processing of usage data in the Navy was held at the U. S. Naval Supply Research and Development Facility, Bayonne, New Jersey, on May 3-6, 1954. The purpose of this group is to act as a continuing advisory agency on matters pertaining to usage data in the Naval Establishment. The group will perform the following functions: (a) assemble information about data that are being collected and processed, (b) make available information regarding the review of such data, (c) make recommendations, as pertinent, to appropriate authority. The current members of this liaison group are: Cdr. George L. Voegeli, ONR; Capt. J. E. Hamilton, GWULRP; LCDR. O. J. Breen, BuSandA; LCDR. A. D. Cox, BuSandA; Mr. R. S. Kelley, Jr., NSRDF; and Mr. J. R. Simpson, NSRDF.

The Economics Research Project at Princeton University, sponsored by the Office of Naval Research, is concerned with the "Mathematical Structure of American-Type Economies," covering investigations on a wide range of problems. Although theoretical, much of the Project

work is more closely related to reality than the usual academic economic theory. The Project is of small size, consisting of Professor Oskar Morgenstern as director, two research associates, three research assistants, and a secretary. An indication of the scope of the topics dealt with is given in the description that follows:

1. The theoretical structure underlying various models of the national economy, rather than the theory of the firm, is studied intensively. For example, the Leontief input-output model represents the economy by a set of linear equations. One of the most crucial problems is the manner in which various sectors of the economy are grouped together. Thus the problem of aggregation is studied both from a formal mathematical and from an empirical point of view. In addition, mathematical properties of input-output systems of equations have been explored. Practical methods for the numerical analysis of these properties have been developed. These properties are relevant for answering questions concerning the stability of the system, its rate of expansion, the accuracy of solution, and many other economic questions. The effects of different methods of handling the transportation and trade sectors of the economy have also been explored, as well as those of handling inventories.

2. Recognizing that models of the national economy are useless if the data are inaccurate and inadequate, the general problem of the accuracy and adequacy of economic observations has been studied as well as the more specific problem of the errors which arise in linear systems due to the data. Closely related to the data problem is that of computation.

Large systems of equations representing the economy can be solved only by modern large-scale computing machinery. The fact that high speed computers are now available is of great significance to economics, and this Project is attempting to assess their impact on the economy and on economic theory. Computations may both facilitate experiments in economics and in some cases may be used as a substitute for experimentation. A considerable number of computations have been performed at the instigation of the Project on various electronic computers, thus providing them with experience in this new field.

3. Three problems which are of direct interest to the Navy and are of general interest to the economy have been dealt with in the past three years. They are inventories, capacity of a supply system, and priorities and follow-ups. The work on inventories has resulted in the publication of the book, The Theory of Inventory Management, which contains an analysis of both civil and military inventory policies. The problem of capacity turned out to be closely interconnected with activity analysis as applied to the theory of the firm. Not only was there a difficult economic problem to be faced in attempting to arrive at a set of satisfactory definitions for the various aspects of capacity, but very real problems of data gathering and information flows were encountered in the examination of the supply depot system. As a natural consequence of this work, the Project began to investigate the properties of priorities and follow-up systems. This work is still in progress. The necessity for close cooperation between the men running the naval supply system and research organizations was brought home in a very forcible manner during all of the above-mentioned work. Theoretical work in this field, to be of any use, must be based on good data and a true understanding of the nature of the problems faced. This can only come about by cooperation between naval and university personnel.

4. The director and one other member of the Project have been closely involved in work on game theory. For example, models of several monopolistic competition structures have

been examined. A general examination of some of the basic ideas concerning market forms and the nature of the determination of market price and distribution of resources is under way. Closely tied in with the new developments of game theory and information theory has been the recent interest in organizations.

Numerous papers have been published in a variety of periodicals. In addition, the principal publications in book form are the following:

O. Morgenstern, On the Accuracy of Economic Observations, Princeton University Press, Princeton, 1950. Second revised and enlarged Edition in preparation.

O. Morgenstern, "Prolegomena to a Theory of Organization," RAND RM-734, 1951.

Shephard, R. W., Cost and Production Functions, Princeton University Press, Princeton, 1953.

Whitin, T. M., The Theory of Inventory Management, Princeton University Press, Princeton, 1953.

O. Morgenstern, ed., (M. Shubik and W. M. Young, authors), "Capacity Analysis of Naval Supply Activities," October 1953 (U.S. Naval Supply Research and Development Facility).

O. Morgenstern, ed., Economic Activity Analysis, John Wiley and Sons, Inc., New York, 1954.

In addition, the following books are in preparation:

Shubik, M., Competition and the Theory of Games.

Peston, M., Aggregation in Pure and Applied Economics.

Zabel, E. J., The Measurement of the Capacity of Economic Systems.

O. Morgenstern and Y. K. Wong, Modern Mathematics for Economists.

* * *

RECENT PUBLICATIONS

DESIGN FOR DECISION. By Irwin D. J. Bross, Macmillan Co., 1953, 276 pp.

This book describes in non-technical language some of the basic ideas underlying statistical decision procedures. Recently developed methods of statistical decision are intended to help one select one course of action from a number of alternative courses of action. Hardly any mathematics is used in the book. Some of the topics discussed in everyday language are: prediction, probability, Bayes rule, risk, sequential procedures, mathematical models, and sampling. The reader who desires to learn the technical aspects of the subject matter, is referred to a list of the leading texts classified according to subject matter, nature of contents, and mathematical level, given at the end of the book.

AN INTRODUCTION TO LINEAR PROGRAMMING. By A. Charnes, W. W. Cooper, and A. Henderson, John Wiley & Sons, Inc., 1953, 74 pp.

This book is composed of two major parts: an introduction and a mathematical development. In the introduction, or Part I, the authors work out a concrete problem in linear programming which is so arranged that the reader need not follow through the entire solution. The reader concerned only with securing some notion of the nature of linear programming should study certain sections indicated in the Preface of the book, while the person interested in a concrete application should add certain additional sections. Part II on the mathematical theory of linear programming, presents a systematic development of the general theory. The presentation, based on a series of lectures, is designed so that a minimum of background in mathematics is required. A bibliography has been appended as a guide to further reading.

A TEXTBOOK OF ECONOMETRICS. By Lawrence R. Klein, Row, Peterson & Co., 1953, 355 pp.

This textbook contains a discussion of most of the standard problems of estimation, computation, and application of econometric models. It considers both highly aggregated structures and multi-sector situations. Thus, the identification problem and the procedures of maximum likelihood are examined in relation to a "global" model of the economy. This is followed by an analysis of a disaggregated model in which account is taken of cross section data as well as time series. A study of production of railway services is presented in detail as an example, and some space is devoted to Leontief's input-output approach. The author then examines various questions of prediction, and goes into more technical matters arising from errors of observation and measurement and other characteristics of the data. The technique of weighted regressions is explained, after which comes a statement on the consequences of varying the length of unit observation periods. The book concludes with an elementary introduction to matrices and determinants.

ECONOMIC CHANGE. By Simon Kuznets, W. W. Norton & Co., Inc., 1953, 333 pp.

This volume consists of eleven essays, nine of which were previously published. Although some of these papers were written over twenty years ago, only minor changes were made; mainly, the deletion of paragraphs no longer pertinent. The papers selected by the author for this volume are the ones which deal with broad problems. All of them are concerned with the movement of economic processes over time. The titles of the essays are as follows: Equilibrium Economics and Business-Cycle Theory; Static and Dynamic Economics; Relation between Capital Goods and Finished Products in the Business Cycle; Schumpeter's Business Cycles; Economic Trends and Business Cycles; National Income and Industrial Structure; National Income and Economic Welfare; International Differences in Income Levels; Retardation of Industrial Growth; Economic Tendencies, Past and Present; Foreign Economic Relations of the United States and their Impact upon the Domestic Economy.

STUDIES IN THE STRUCTURE OF THE AMERICAN ECONOMY. By Wassily Leontief and others, Oxford University Press, 1953, 561 pp.

This volume presents some of the results of the Harvard Economic Research Project on the Structure of the American Economy. It consists of twelve chapters grouped into five parts and three appendices. The theoretical framework of this volume is the input-output approach made popular by the principal author. The titles of the five parts are: Static and Dynamic Theory, The Extension of Input-Output Techniques to Interregional Analysis, The Capital Structure of the American Economy, Explorations in the Use of Technological Data, Consumption and Final Demand. The three appendices are entitled: Summary of 1939 Capital Structure of American Industries, Comparisons of Input-Output Classifications, and Estimation of Price Elasticities from Budget Studies.

LOGISTICAL SUPPORT OF THE ARMIES, VOL. I. By Roland G. Ruppenthal, U. S. Government Printing Office, 1953, 616 pp.

This volume, another in the series U. S. Army in World War II, is the third to be published in the subseries The European Theater of Operations. This volume describes how the United States forces were built up in the United Kingdom for the invasion of 1944 and how they were supplied during the first three months of operation on the European Continent. The author relates the problems of logistic support to tactical plans and operations. The story of the procurement, movement, and distribution of supplies and manpower is told largely from the standpoint of the agency responsible for the support of United States forces; however, the focus throughout is on the influence which logistic support or lack of it had on the planning and conduct of combat operations by the field armies. Considerable space is given to the discussions of theater command and organization. The volume contains many charts, maps, and illustrations. In the second volume, the story will be continued to the end of hostilities in Europe in May, 1945.

COST AND PRODUCTION FUNCTIONS. By Ronald W. Shephard, Princeton University Press, 1953, 104 pp.

This book is a rather mathematical treatment of the economic theory of production. In the first five sections the author attempts to formulate an integrated theory of cost and production functions with the goal of extending the basis of G. C. Evans' static and dynamic economic analyses. The next four sections are directed to the development of an aggregated form of the classical problem of measurement of a production function in preference to a microscopic study like that of the activity analysis model of production. However, available technological information as to the structure of production are incorporated into the mathematical form of the production function. In the tenth and last section the author reformulates Evans' dynamic analysis of monopoly in terms of a cost function which allows for anticipations by the entrepreneur of changes in cost associated with prospective variations in the prices of the factors of production.

MATHEMATICS AND STATISTICS FOR ECONOMISTS. By Gerhard Tintner, Rinehart & Co., Inc., 1953, 363 pp.

This textbook is intended for the American student of economics who desires to acquire the minimum training in mathematics and statistics needed for a study of econometrics. Most of the illustrative examples and problems are taken from the field of economics. The book does not assume that the reader has had any previous college training in mathematics or statistics. Intuitive proofs are frequently substituted for mathematically rigorous ones because the book is not intended for the student of mathematics.

DEMAND ANALYSIS. By Herman Wold in association with Lars Juréen, John Wiley & Sons, Inc., 1953, 358 pp.

This volume consists of five parts, the first of which starts out with a discussion of demand analysis methods, including illustrations of the methods by empirical studies made by the authors on the structure of consumer demand in Sweden. Least-squares regression analysis, the principal method of demand analysis, is discussed in some detail. Part II is concerned with Pareto's theory of preference fields with special attention paid to interrelations between demand elasticities. Part III gives a brief survey of the theory of stationary processes. Part IV presents a more theoretical treatment of least-squares regression analysis than that presented in Part I. The final part contains the empirical findings of the authors obtained from studies made by the authors on the demand structure in Sweden. The findings were derived from data obtained from family budget surveys and market statistics.

* * *

☆ U. S. GOVERNMENT PRINTING OFFICE : 1954 O —318460